



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number 126106**

**TO: Ginny Portner**  
**Location: REM/3B02/3C18**  
**Art Unit: 1645**  
**Tuesday, July 06, 2004**

**Case Serial Number: 09/732091**

**From: Toby Port**  
**Location: Biotech-Chem Library**  
**Remsen 1A59**  
**Phone: 571-272-2523**

**toby.port@uspto.gov**

### **Search Notes**

Dear Examiner Portner,

Here are the results of your search.  
Please feel free to contact me if you have any questions.

Toby Port

STIC-Biotech/ChemLib

196106

From: Portner, Ginny  
Sent: Tuesday, June 29, 2004 6:07 PM  
To: STIC-Biotech/ChemLib  
Subject: RE: 09/732,091

try 09/732,091

*Ginny Portner*  
Remsen Building  
Art Unit 1645  
Room E03, B02  
(571) 272-0862

-----Original Message-----

From: STIC-Biotech/ChemLib  
Sent: Tuesday, June 29, 2004 4:23 PM  
To: Portner, Ginny  
Subject: RE: 09/732,081

There is no valid CRF for this Serial Number. Please provide us with another Serial Number from the Parent Application.  
LEONARD 22520

-----Original Message-----

From: Portner, Ginny  
Sent: Tuesday, June 29, 2004 4:21 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/732,081  
Importance: High

Please search SEQ ID NO 3 and 4 with respect to polypeptides/proteins. The nucleic acid should be back translated to the polypeptide. Thanks.

*Ginny Portner*  
Remsen Building  
Art Unit 1645  
Room E03, B02  
(571) 272-0862

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

CC Helicobacter sp. HP30 protein  
 XX Sequence 253 AA;  
 SQ  
 Query Match 100.0%; Score 1279; DB 5; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-118;  
 Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAYKYDRDLFLKQLESSDLDLFEVLVFGKDGKRRNEKLTSSIEYKRGHGGDYAKYAE 60  
 Db 1 MAYKYDRDLFLKQLESSDLDLFEVLVFGKDGKRRNEKLTSSIEYKRGHGGDYAKYAE 60  
 QY 61 IAEELQYGSNSPASFIFKGEGLVYKEILCDVCDKLVNKNKTTTLLIEQNMLSKILERS 120  
 Db 61 IAEELQYGSNSPASFIFKGEGLVYKEILCDVCDKLVNKNKTTTLLIEQNMLSKILERS 120  
 QY 121 LEEMDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFMGGFKSYQLAVIVANAVAKTIL 180  
 Db 121 LEEMDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFMGGFKSYQLAVIVANAVAKTIL 180  
 QY 181 GRGLSLAGNOVLTRTSLFSLTGPVGMIIITGVMTAIDAGPAYRVVTIPACIVVATLRLKTOO 240  
 Db 181 GRGLSLAGNOVLTRTSLFSLTGPVGMIIITGVMTAIDAGPAYRVVTIPACIVVATLRLKTOO 240  
 QY 241 ANGDKKSLQIESI 253  
 Db 241 ANGDKKSLQIESI 253  
 13-DEC-2002 (first entry)  
 Helicobacter sp. PQE/HP30 protein.  
 HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
 type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;  
 low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
 immunostimulant; cytostatic; peptic ulcer.  
 Helicobacter sp.  
 Key Location/Qualifiers  
 FT Misc-difference 12 /note= "Encoded by TTC"  
 FT W0200251237-A2.  
 PN 04-JUL-2002.  
 PD 07-DEC-2001; 2001WO-US048392.  
 PF 07-DEC-2000; 2000US-00732091.  
 PR (ANTE-) ANTEX BIOLOGICS INC.  
 PA Tian J, Walker R, Jackson WJ;  
 PI WPI; 2002-666854/71.  
 PP N-PSDB; AAD44535.  
 DR Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
 PT the proteins, useful as vaccines for raising immune response in animals.  
 PS Disclosure; Page 122-123; 127pp; English.  
 XX The invention relates to Helicobacter HP30 or HP56 polypeptide and  
 CC peptides derived from them. The invention is useful for producing an  
 CC immune response. It is useful for preventing, treating or ameliorating a

CC disorder or disease associated with infection of an animal with  
 CC Helicobacter. Pharmaceutical composition and vaccines comprising the  
 CC sequences of the invention is useful for treating type B gastritis,  
 CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B  
 CC cell lymphoma. The invention is useful as reagents for clinical or  
 CC medical diagnosis of Helicobacter infections and for scientific research  
 CC on the properties of pathogenicity, virulence and infectivity of  
 CC Helicobacter, as well as host defense mechanisms. The present sequence is  
 CC Helicobacter sp. PQE/HP30 protein  
 XX Sequence 265 AA;  
 SQ  
 Query Match 100.0%; Score 1279; DB 5; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-118;  
 Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAYKYDRDLFLKQLESSDLDLFEVLVFGKDGKRRNEKLTSSIEYKRGHGGDYAKYAE 60  
 Db 13 MAYKYDRDLFLKQLESSDLDLFEVLVFGKDGKRRNEKLTSSIEYKRGHGGDYAKYAE 72  
 QY 61 IAEELQYGSNSPASFIFKGEGLVYKEILCDVCDKLVNKNKTTTLLIEQNMLSKILERS 120  
 Db 73 IAEELQYGSNSPASFIFKGEGLVYKEILCDVCDKLVNKNKTTTLLIEQNMLSKILERS 132  
 QY 121 LEEMDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFMGGFKSYQLAVIVANAVAKTIL 180  
 Db 133 LEEMDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFMGGFKSYQLAVIVANAVAKTIL 192  
 QY 181 GRGLSLAGNOVLTRTSLFSLTGPVGMIIITGVMTAIDAGPAYRVVTIPACIVVATLRLKTOO 240  
 Db 193 GRGLSLAGNOVLTRTSLFSLTGPVGMIIITGVMTAIDAGPAYRVVTIPACIVVATLRLKTOO 252  
 QY 241 ANGDKKSLQIESI 253  
 Db 253 ANGDKKSLQIESI 265  
 RESULT 3  
 AAW20486  
 ID AAW20486 standard; protein; 253 AA.  
 XX AC AAW20486;  
 XX 29-JUL-1997 (first entry)  
 DT H. pylori cytoplasmic protein, 4095342.aa.  
 DE Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 XX Helicobacter pylori.  
 OS WO9640893-A1.  
 XX 19-DEC-1996.  
 PD 05-JUN-1996; 96WO-US009122.  
 PF 07-JUN-1995; 95US-00487032.  
 PR 01-APR-1996; 96US-00630405.  
 XX (ASTR ) ASTRA AB.  
 PA Smith D, Berglindh OT, Mellgaard BL;  
 XX WPI; 1997-052306/05.  
 DR N-PSDB; AAT67811.  
 DR Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 PT useful for vaccines to treat or prevent H. pylori infection, and to  
 PT detect Helicobacter.  
 XX

PS Claim 61; Page 651; 1481pp; English.

XX The present sequence is a H. pylori cytoplasmic protein. The protein may  
 CC be used in a vaccine to prevent or treat H. pylori infection or to  
 CC identify H. pylori polypeptide binding compounds, useful as potential H.  
 CC pylori life cycle activators or inhibitors. The genomic sequence of H.  
 CC pylori (ATCC 55679) was determined from overlapping contigs generated by  
 CC mechanically shearing the bacterial DNA. The sequences were analysed for  
 CC ORF of at least 180 nucleotides, and the predicted coding regions defined  
 CC by computer evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF were  
 CC analysed for significant homology to other known or exported membrane  
 CC proteins. Having identified and determined the sequences of interest,  
 CC particular regions can be isolated from H. pylori by PCR amplification  
 CC for recombinant polypeptide production, e.g. in E. coli hosts

XX Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;  
 Best Local Similarity 99.2%; Pred. No. 1.4e-117;  
 Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAYKYRDLFLKQLSSDLDLFEVLVFGKDGKRNKLTSSIEYKRHGDYAKYAE 60  
 DB 1 MAYKYRDLFLKQLSSDLDLFEVLVFGKDGKRNKLTSSIEYKRHGDYAKYAE 60  
 QY 61 IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKNKTETTLIEQNMLSKILERS 120  
 DB 61 IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKNKTETTLIEQNMLSKILERS 120  
 QY 121 LEEMDDDEVKEMCDELISIKNTDNLNRQALSAAATLTFKMGGPKSYQLAVIVANAVAKTIL 180  
 DB 121 LEEMDDDEVKEMCDELISIKNTDNLNRQALSAAATLTFKMGGPKSYQLAVIVANAVAKTIL 180  
 QY 181 GRGLSLAGNOVLTRTSLFSLTGPVGMWITGVWTAIDAGPAYRVITIPACIVVATLRKLTQQ 240  
 DB 181 GRGLSLAGNOVLTRTSLFSLTGPVGMWITGVWTAIDAGPAYRVITIPACIVVATLRKLTQQ 240  
 QY 241 ANGDKKSLQIESI 253  
 DB 241 ANEDKKSLQIESV 253

RESULT 4  
 AAW24673  
 ID AAW24673 standard; protein; 253 AA.

AC AAW24673  
 DT 12-AUG-1997 (first entry)

XX H. pylori cytoplasmic protein, 4095342.aa.  
 XX Transmembrane; cytoplasmic; cell envelope; flagella; transport; secreted;  
 KW periplasmic; chronic gastritis; duodenal ulcer disease; activator;  
 KW inhibitor; bacterial life cycle; vaccine; immunise; detection; antisense;  
 XX inhibition.  
 XX Helicobacter pylori.  
 XX WO9719098-A1.  
 XX 29-MAY-1997.  
 XX 15-NOV-1996; 96WO-US018542.  
 XX 17-NOV-1995; 95US-00561469.  
 XX (ASTR ) ASTRA AB.  
 XX Smith DH;  
 XX WPI; 1997-298052/27.

DR N-PSDB; AAT77491.  
 XX Helicobacter pylori nucleic acid sequences and related proteins - used  
 PT for diagnostics and therapeutics.  
 XX Claim 18; Page 184; 235pp; English.

XX This sequence represents an H. pylori cytoplasmic protein. Helicobacter  
 CC pylori has been strongly linked to chronic gastritis and duodenal ulcer  
 CC disease. The nucleic acid sequences of the invention are used to evaluate  
 CC compounds, especially activators or inhibitors of bacterial life cycle,  
 CC for the ability to bind an H. pylori nucleic acid sequence. The nucleic  
 CC acid sequences, and corresponding proteins, are also useful for  
 CC generating vaccines for immunising subjects against H. pylori or for use  
 CC in detecting the presence of Helicobacter species in a sample. Antisense  
 CC nucleic acid sequences of these sequences are used to inhibit expression  
 CC of a gene from Helicobacter species. H. pylori whole genomic DNA was  
 CC isolated and nebulised to a median size of 2000 bp. Purified DNA  
 CC fragments were blunt-ended and ligated to unique BstXI-linker adapters in  
 CC 100-1000 fold molar excess. These linkers are complementary to the BstXI-  
 CC cut PMX vectors, while the overhang is not self-complementary. Therefore  
 CC the linkers will not concatemerise nor will the cut vector re-ligate  
 CC itself easily. The linker-adaptor inserts were ligated to each of the 20  
 CC PMX vectors to construct a series of shotgun subclone libraries. The  
 CC purified DNA samples were then sequenced. Note: The ORF/protein reference  
 CC number for this sequence was obtained from the related specification,  
 CC WO9640893

XX Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;  
 Best Local Similarity 99.2%; Pred. No. 1.4e-117;  
 Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAYKYRDLFLKQLSSDLDLFEVLVFGKDGKRNKLTSSIEYKRHGDYAKYAE 60  
 DB 1 MAYKYRDLFLKQLSSDLDLFEVLVFGKDGKRNKLTSSIEYKRHGDYAKYAE 60  
 QY 61 IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKNKTETTLIEQNMLSKILERS 120  
 DB 61 IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKNKTETTLIEQNMLSKILERS 120  
 QY 121 LEEMDDDEVKEMCDELISIKNTDNLNRQALSAAATLTFKMGGPKSYQLAVIVANAVAKTIL 180  
 DB 121 LEEMDDDEVKEMCDELISIKNTDNLNRQALSAAATLTFKMGGPKSYQLAVIVANAVAKTIL 180  
 QY 181 GRGLSLAGNOVLTRTSLFSLTGPVGMWITGVWTAIDAGPAYRVITIPACIVVATLRKLTQQ 240  
 DB 181 GRGLSLAGNOVLTRTSLFSLTGPVGMWITGVWTAIDAGPAYRVITIPACIVVATLRKLTQQ 240  
 QY 241 ANGDKKSLQIESI 253  
 DB 241 ANEDKKSLQIESV 253

RESULT 5  
 AAW20866  
 ID AAW20866 standard; protein; 256 AA.

XX AAW20866;  
 XX 18-JUL-1997 (first entry)

XX H. pylori cytoplasmic protein, 12ge20305orf30.  
 XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis.  
 XX Helicobacter pylori.  
 XX WO9640893-A1.



QY 1 ATGGCATACAAATATGATAGAGACTTGGAAATTTTAAAGCAATTCGAATCTAGTATT 60  
 Db 13 MetAlaTyrlsTyAspArgAspLeuGluPheLeuGlnLeuGluSerSerAspLeu 32  
 QY 61 TTGGATTGTTGGAGTCTGTTGTTTGGTAAAGACGCGGAAAGACACAAATGAA 120  
 Db 33 LeuAspLeuPheGluValLeuValPheGlyAspGlyGluValArgHisAsnGluLys 52  
 QY 121 CTGACCACTCCATAGATACAAAGGATGCGCATGATTAACGTAATACGCAAGAA 180  
 Db 53 LeuThrSerSerlleGluTyrlsArgHisGlyAspAspTyrlsAlaLysTyrlsAlaGluArg 72  
 QY 181 ATCCCTCAAGAGTTGCAATCTATGCGAGCAATAGTTTTCGAGCTTCATTAAGGCGAA 240  
 Db 73 IleAlaGluGluLeuGlnTyrlsTyrlsGlySerAsnSerPheAlaSerPheIleLysGlyGlu 92  
 QY 241 GGAGCTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAGGTCATTAACAAC 300  
 Db 93 GlyValLeuTyrlsGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrlsAsn 112  
 QY 301 AAGAAACTGAAAGCACTTAATTAACAAACATGCTTCTTAAATCTTAGAAAGAGT 360  
 Db 113 LysLysThrGluThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 132  
 QY 361 TTGGAGAAATGATGATGAAGAAGTCAAGAAATGTGCGATGAATTCATTAACAAC 420  
 Db 133 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerlleLysAsn 152  
 QY 421 ACGGACAAATTAACACAGCAAGCTTAAGCGCGCGACTTTAACGCTGTTTAAATGGGG 480  
 Db 153 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 172  
 QY 481 GGTTTTAAATCTTATCAATAGCTGTCATTTGCGAATGCGTCCGCAAAACCATCTTA 540  
 Db 173 GlyPheLysSerTyrlsGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 192  
 QY 541 GGGCGTGGTTATCGCTTCGGGCAATCAAGTCTTACAAAGCACTCTGAGCTTTTAAACA 600  
 Db 193 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 212  
 QY 601 GTCCTGTTGGCTGGATCATTACAGCGGTATGACAGCGATTCGATTCAGCGCGCGCT 660  
 Db 213 GlyProValGlyTrpIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 232  
 QY 661 TATAGGTAACCATACCGCATGCTGTTGCTGCTTACAGCACTTACGCTTAAACACAGCAA 720  
 Db 233 TyrArgValThrIleProAlaCysIleValAlaThrLeuArgLeuLysThrGlnGln 252  
 QY 721 GCCAATGGAGATAGAAGTCGTTGCAATAGATCAAT 759  
 Db 253 AlaAsnGlyAspLysSerLeuGlnIleGluSerlle 265

## RESULT 3

AAW20486

ID AAW20486 standard; protein; 253 AA.

XX AC AAW20486;

XX XX

DT 29-JUL-1997 (first entry)

XX XX

DE XX

H. pylori cytoplasmic protein, 4095342.aa.

XX XX

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

XX XX

OS Helicobacter pylori.

XX XX

W09640893-A1.

XX XX

PD 19-DEC-1996.

XX XX

FF 06-JUN-1996; 96WO-US009122.

XX 07-JUN-1995; 95US-00487032.  
 PR 01-APR-1996; 96US-00630405.  
 XX (ASTR ) ASTRA AB.  
 XX Smith D, Berglindh OT, Melligaard BL;  
 PI WPI; 1997-052306/05.  
 XX N-PSDB; AAT67811.  
 DR XX

PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 useful for vaccines to treat or prevent H. pylori infection, and to  
 detect Helicobacter.

PS Claim 61; Page 651; 1481pp; English.

XX The present sequence is a H. pylori cytoplasmic protein. The protein may  
 be used in a vaccine to prevent or treat H. pylori infection or to  
 identify H. pylori polypeptide binding compounds, useful as potential H.  
 pylori life cycle activators or inhibitors. The genomic sequence of H.  
 pylori (ATCC 55679) was determined from overlapping contigs generated by  
 mechanically shearing the bacterial DNA. The sequences were analysed for  
 ORF of at least 180 nucleotides, and the predicted coding regions defined  
 by computer evaluation. To identify likely H. pylori antigens for vaccine  
 development, the amino acid sequences predicted from various ORF were  
 analysed for significant homology to other known or exported membran  
 proteins. Having identified and determined the sequences of interest,  
 particular regions can be isolated from H. pylori by PCR amplification  
 for recombinant polypeptide production, e.g. in E. coli hosts

XX Sequence 253 AA.

## Alignment Scores:

Pred. No.: 6,94e-134 Length: 253  
 Score: 1270.00 Matches: 251  
 Percent Similarity: 99.60% Conservatve: 1  
 Best Local Similarity: 99.21% Mismatches: 1  
 Query Match: 94.14% Indels: 0  
 DB: 2 Gaps: 0

US-09-732-091-3 (1-759) x AAW20486 (1-253)

QY 1 ATGGCATACAAATATGATAGAGACTTGGAAATTTTAAAGCAATTCGAATCTAGTATT 60  
 Db 1 MetAlaTyrlsTyAspArgAspLeuGluPheLeuGlnLeuGluSerSerAspLeu 20  
 QY 61 TTGGATTGTTGGAGTCTGTTGTTTGGTAAAGACGCGGAAAGACACAAATGAA 120  
 Db 21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40  
 QY 121 CTGACCACTCCATAGATACAAAGGATGCGCATGATTCGCTTAAATACGCAAGAA 180  
 Db 41 LeuThrSerSerlleGluTyrlsArgHisGlyAspAspTyrlsAlaLysTyrlsAlaGluArg 60  
 QY 181 ATCCCTCAAGAGTTGCAATCTATGCGAGCAATAGTTTTCGAGCTTCATTAAGGCGAA 240  
 Db 61 IleAlaGluGluLeuGlnTyrlsTyrlsGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80  
 QY 241 GGAGCTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAGGTCATTAACAAC 300  
 Db 81 GlyValLeuTyrlsGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrlsAsn 100  
 QY 301 AAGAAACTGAAAGCACTTAATTAACAAACATGCTTCTTAAATCTTAGAAAGAGT 360  
 Db 101 LysLysThrGluThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120  
 QY 361 TTGGAGAAATGATGATGAAGAAGTCAAGAAATGTGCGATGAATTCATTAACAAC 420  
 Db 121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerlleLysAsn 140  
 QY 421 ACGGACAAATTAACACAGCAAGCTTTAAAGCGCGCGACTTTAACGCTGTTTAAATGGGG 480

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: July 5, 2004, 03:11:28 ; Search time 34 Seconds  
(without alignments)  
2316.318 Million cell updates/sec

Title: US-09-732-091-4

Perfect score: 1279  
Sequence: 1 MAYKYDRLEFLKQLESSDL.....LRLXTQANGDKKSLQIESI 253

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/2/pubaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1279	100.0	253	9 US-09-732-091-4	Sequence 4, Appl
2	1279	100.0	265	9 US-09-732-091-44	Sequence 44, Appl
3	1270	99.3	253	12 US-10-335-977-9162	Sequence 9162, Ap
4	1270	99.3	253	12 US-10-335-977-9163	Sequence 9163, Ap
5	1270	99.3	256	12 US-10-335-977-9164	Sequence 9164, Ap
6	722	56.5	248	12 US-10-335-977-7699	Sequence 7699, Ap
7	716	56.0	237	12 US-10-335-977-7698	Sequence 7698, Ap
8	473.5	36.9	155	10 US-09-882-227-414	Sequence 4-4, App
9	252	19.7	49	9 US-09-732-091-20	Sequence 20, Appl
10	198	15.5	38	9 US-09-732-091-17	Sequence 17, Appl
11	194	15.2	41	9 US-09-732-091-19	Sequence 19, Appl
12	152	11.9	30	9 US-09-732-091-18	Sequence 18, Appl
13	150	11.7	30	9 US-09-732-091-16	Sequence 16, Appl
14	98.5	7.7	1009	12 US-10-282-122A-43832	Sequence 43832, A
15	97.5	7.6	815	16 US-10-437-963-191043	Sequence 191043,

15	97	7.6	916	12	US-10-282-122A-76490	Sequence 76490, A
17	96.5	7.5	373	15	US-10-369-493-42	Sequence 42, Appl
18	96.5	7.5	701	14	US-10-032-585-7400	Sequence 7400, Ap
19	96.5	7.5	1102	14	US-10-156-761-14395	Sequence 14395, A
20	95	7.4	517	12	US-10-282-122A-54543	Sequence 54543, A
21	95	7.4	2125	9	US-09-919-172-29	Sequence 29, Appl
22	95	7.4	2649	14	US-10-205-219-169	Sequence 169, Appl
23	95	7.4	2649	15	US-10-341-434-220	Sequence 220, App
24	95	7.4	2649	15	US-10-341-434-230	Sequence 230, App
25	93.5	7.3	1196	12	US-10-282-122A-52737	Sequence 52737, A
26	93	7.3	610	12	US-10-282-122A-47217	Sequence 47217, A
27	92.5	7.2	430	15	US-10-369-493-5229	Sequence 5229, Ap
28	92.5	7.2	996	9	US-09-815-242-5251	Sequence 5251, Ap
29	92.5	7.2	1009	9	US-09-815-242-12141	Sequence 12141, A
30	92	7.2	889	16	US-10-437-963-197045	Sequence 197045,
31	91.5	7.2	241	12	US-10-412-699B-682	Sequence 682, App
32	91.5	7.2	1847	15	US-10-369-493-1075	Sequence 1075, Ap
33	90.5	7.1	659	12	US-10-424-599-222059	Sequence 222059,
34	90.5	7.1	1163	15	US-10-452-024-107	Sequence 107, App
35	90	7.0	573	12	US-10-425-114-43207	Sequence 43207, A
36	90	7.0	744	16	US-10-437-963-122954	Sequence 122954,
37	90	7.0	820	12	US-10-221-278-590	Sequence 590, App
38	90	7.0	820	15	US-10-291-172-590	Sequence 590, App
39	90	7.0	1089	14	US-10-032-585-7664	Sequence 7664, Ap
40	89.5	7.0	281	15	US-10-289-782-749	Sequence 749, Appl
41	89.5	7.0	560	14	US-10-253-904-50	Sequence 50, Appl
42	89.5	7.0	869	12	US-10-282-122A-58561	Sequence 58561, A
43	89	7.0	384	12	US-10-260-708-60	Sequence 60, Appl
44	88.5	6.9	482	12	US-10-424-599-199420	Sequence 199420,
45	88.5	6.9	552	9	US-09-817-764-4	Sequence 4, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-732-091-4  
; Sequence 4, Application US/09732091  
; Patent No. US20020107368A1  
; GENERAL INFORMATION:  
; APPLICANT: Tian, Jing-Hui  
; APPLICANT: Walker, Richard I.  
; APPLICANT: Jackson, W. James  
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses  
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses  
; FILE REFERENCE: 7969-088  
; CURRENT APPLICATION NUMBER: US/09/732,091  
; CURRENT FILING DATE: 2000-12-07  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 253  
; TYPE: PRT  
; ORGANISM: Helicobacter sp.  
US-09-732-091-4

Query Match	100.0%	Score 1279;	DB 9;	Length 253;
Best Local Similarity	100.0%	Pred. No. 4.8e-115;		
Matches 253;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAYKYDRLEFLKQLESSDL	1	MAYKYDRLEFLKQLESSDL
Db	1	MAYKYDRLEFLKQLESSDL	1	MAYKYDRLEFLKQLESSDL
QY	61	IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKKTKTTTLEQNMLSKILERS	120	IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKKTKTTTLEQNMLSKILERS
Db	61	IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKKTKTTTLEQNMLSKILERS	120	IAEELQYGSNSFASFIKGEVLYKEILCDVCDKLVNKKTKTTTLEQNMLSKILERS
QY	121	LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTFLKMGFKSYQLAVIVANAVAKTLL	180	LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTFLKMGFKSYQLAVIVANAVAKTLL
Db	121	LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTFLKMGFKSYQLAVIVANAVAKTLL	180	LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTFLKMGFKSYQLAVIVANAVAKTLL
QY	181	GRGLSIAGNQLVTRTSLFTGPGVMIITGWTWTAIDAGPAYRVVTIPACIVVATLRLKTKQ	240	GRGLSIAGNQLVTRTSLFTGPGVMIITGWTWTAIDAGPAYRVVTIPACIVVATLRLKTKQ



COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 03/993,002  
FILING DATE: 17-Dec-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 9163:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...253  
SEQUENCE DESCRIPTION: SEQ ID NO: 9163:  
US-10-335-977-9163

Query Match 99.3%; Score 1270; DB 12; Length 253;  
Best Local Similarity 99.2%; Pred. No. 3.6e-114;  
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAYKDRDLFLKQLLESSDLLDFVLVFGKDGKGRHNEKLTSSIEYKRGHGDYAKYAE 60  
DB 1 MAYKDRDLFLKQLLESSDLLDFVLVFGKDGKGRHNEKLTSSIEYKRGHGDYAKYAE 60  
QY 61 LAELQYGSNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
DB 61 LAELQYGSNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
QY 121 LEEMDDERVKEMCDLSIKNTNLNRQALSAATLTLFKMGGPKSYQLAVIVANAVAKTIL 180  
DB 121 LEEMDDERVKEMCDLSIKNTNLNRQALSAATLTLFKMGGPKSYQLAVIVANAVAKTIL 180  
QY 181 GRGLSLAGNQLVTRLTSLFTGPGVMIITGVWTAIDTAGPAYRVTIPACIVWATLRLKTOQ 240  
DB 181 GRGLSLAGNQLVTRLTSLFTGPGVMIITGVWTAIDTAGPAYRVTIPACIVWATLRLKTOQ 240  
QY 241 ANGDKKSQIESI 253  
DB 241 ANEDKKSQIESV 253

RESULT 5  
US-10-335-977-9164  
Sequence 9164, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts

COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-Dec-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 9164:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...256  
SEQUENCE DESCRIPTION: SEQ ID NO: 9164:  
US-10-335-977-9164

Query Match 99.3%; Score 1270; DB 12; Length 256;  
Best Local Similarity 99.2%; Pred. No. 3.6e-114;  
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAYKDRDLFLKQLLESSDLLDFVLVFGKDGKGRHNEKLTSSIEYKRGHGDYAKYAE 60  
DB 4 MAYKDRDLFLKQLLESSDLLDFVLVFGKDGKGRHNEKLTSSIEYKRGHGDYAKYAE 63  
QY 61 LAELQYGSNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
DB 64 LAELQYGSNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 123  
QY 121 LEEMDDERVKEMCDLSIKNTNLNRQALSAATLTLFKMGGPKSYQLAVIVANAVAKTIL 180  
DB 124 LEEMDDERVKEMCDLSIKNTNLNRQALSAATLTLFKMGGPKSYQLAVIVANAVAKTIL 183  
QY 181 GRGLSLAGNQLVTRLTSLFTGPGVMIITGVWTAIDTAGPAYRVTIPACIVWATLRLKTOQ 240  
DB 184 GRGLSLAGNQLVTRLTSLFTGPGVMIITGVWTAIDTAGPAYRVTIPACIVWATLRLKTOQ 243  
QY 241 ANGDKKSQIESI 253  
DB 244 ANEDKKSQIESV 256

RESULT 6  
US-10-335-977-7699  
Sequence 7699, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street

CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-Dec-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 742-4214  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 7699:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 248 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...248  
SEQUENCE DESCRIPTION: SEQ ID NO: 7699:  
US-10-335-977-7699

Query Match 56.5%; Score 722; DB 12; Length 248;  
Best Local Similarity 60.3%; Pred. No. 2.8e-61;  
Matches 141; Conservative 36; Mismatches 53; Indels 4; Gaps 2;  
QY 8 DLEFLKQLESSDLDLFEVLVFGKGEKRNHNEKLTSSIEYKRGHDDYAKYAEELQY 67  
Db 11 DLEFLKRLSSDLKDLFDALVDEDTLRNNEELTSLTEYQYGHYAKYPRRIAEELQ 70  
QY 68 YGNSFASFKGEVLKYLCDVCDKLVKNYKNTTTLIEQNMLSKILERSLEEMDDE 127  
Db 71 YGNSFANFRDEGVLYKEILCDACDHLDINYNERSATSLIEQNMLSKILKDSLEKMSGR 130  
QY 128 EVKEMCDELSIKNTNL---NRQALSAATLTLPKMGFKSYQLAVIVANAVAKTILGRGL 184  
Db 131 EIKELCDGLGMPNIDKVIENKQVLIAVLTLPKAGGSHSYALAVAVADAMVQTLGHGL 190  
QY 185 -SIAGNQLVTRTSLFTGPGVWIIITGVTATIDAGPAYRVITPACIVVATLRK 237  
Db 191 SSVGVKVALKXKTLIDILAGPIGWITGALVSINLAGPAYRVITPACIVVATLRKK 244

RESULT 7  
US-10-335-977-7698  
Sequence 7698, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts

COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-Dec-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 742-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 7698:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 237 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...237  
SEQUENCE DESCRIPTION: SEQ ID NO: 7698:  
US-10-335-977-7698

Query Match 56.0%; Score 716; DB 12; Length 237;  
Best Local Similarity 60.1%; Pred. No. 1e-60;  
Matches 140; Conservative 36; Mismatches 53; Indels 4; Gaps 2;  
QY 9 DLEFLKQLESSDLDLFEVLVFGKGEKRNHNEKLTSSIEYKRGHDDYAKYAEELQY 68  
Db 1 DLEFLKRLSSDLKDLFDALVDEDTLRNNEELTSLTEYQYGHYAKYPRRIAEELQY 60  
QY 69 GNSFASFKGEVLKYLCDVCDKLVKNYKNTTTLIEQNMLSKILERSLEEMDDE 128  
Db 61 GNSFANFRDEGVLYKEILCDACDHLDINYNERSATSLIEQNMLSKILKDSLEKMSGR 120  
QY 129 VKEMCDELSIKNTNL---NRQALSAATLTLPKMGFKSYQLAVIVANAVAKTILGRGL- 184  
Db 121 IKELCDGLGMPNIDKVIENKQVLIAVLTLPKAGGSHSYALAVAVADAMVQTLGHGLS 180  
QY 185 SIAGNQLVTRTSLFTGPGVWIIITGVTATIDAGPAYRVITPACIVVATLRK 237  
Db 181 SSVGVKVALKXKTLIDILAGPIGWITGALVSINLAGPAYRVITPACIVVATLRKK 233

RESULT 8  
US-09-882-227-414  
Sequence 414, Application US/0988227  
Publication No. US20030158396A1  
GENERAL INFORMATION:  
APPLICANT: Kleanthous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Charles  
APPLICANT: Tomb, Jean-Francois  
TITLE OF INVENTION: Identification of Polynucleotides  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts

```
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-414

Query Match      36.9%; Score 472.5; DB 10; Length 155;
Best Local Similarity 63.3%; Pred. No. 1.8e-37;
Matches 93; Conservative 22; Mismatches 29; Indels 3; Gaps 1;

QY 38 NEKLTSSIEYKRGDDYAKYAEIRIAEELQYGGNSFASFIKGGVLYKEILCDVCDKLV 97
Db 2 NEDLTNSTYKRYGHDYAKYPRIAEELQHYGNSFANFRDEGVLYKSIKILCDACDHLKV 61

QY 98 NYNKTETTLIQNNLSKILERSLEEMDDDEEVKEMCDELSIKNTNML--NRQALSAATL 154
Db 62 NYNEESATSLIQNNLSKLLKDSLEKMSRREIKELCNELGWTNIDKVGINKQVLIATSL 121

QY 155 TLFKMGGFKSYQLAVIVANAVAKTILG 181
Db 122 TLFKAGGSHSYALAVSVADAVRQTILG 148

RESULT 9
US-09-732-091-20
; Sequence 20, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteits, gene sequences and uses
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-20

Query Match      19.7%; Score 252; DB 9; Length 49;
Best Local Similarity 100.0%; Pred. No. 6.9e-17;
Matches 49; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 203 VGIITGWTAIDIASPAYRVTIPACIVVATLRLKTOQANGDKKSLQIE 251
Db 1 VGIITGWTAIDIASPAYRVTIPACIVVATLRLKTOQANGDKKSLQIE 49

RESULT 10
US-09-732-091-17
; Sequence 17, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteits, gene sequences and uses
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-17

Query Match      15.5%; Score 198; DB 9; Length 38;
Best Local Similarity 100.0%; Pred. No. 7.9e-12;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 DVAKYAEIRIAEELQYGGNSFASFIKGGVLYKEILCD 90
Db 1 DVAKYAEIRIAEELQYGGNSFASFIKGGVLYKEILCD 38

RESULT 11
US-09-732-091-19
; Sequence 19, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-19

Query Match      15.2%; Score 194; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS 185
Db 1 NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS 41

RESULT 12
US-09-732-091-18
; Sequence 18, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-18

Query Match      11.9%; Score 152; DB 9; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 LEEMDDEEVKEMCDELSIKNTDNLNRQALS 150
Db 1 LEEMDDEEVKEMCDELSIKNTDNLNRQALS 30
```



Db 635 RELEALVKQLQMSSEARKALEADLDEATKSLDEMRALS 674

Search completed: July 5, 2004, 03:13:58  
Jcb time : 34 secs



Db 2 NEULTNSTYKRYGHDVAKYPRRIAELOHYGNGSFANFDEGVLKYLCDACDHLKV 61  
QY 98 NYNKKTTTLEQNMKSKILERSLEEMDDDEVKEMCDLSIKNTDNL---NRQALSATL 154  
Db 62 NYNEESATSIQNMKSKILKCDLSKMSREIKELCNELGNTNIDKVGEMKQVLIASL 121  
QY 155 TLFKMGGFKSYQLAVIVANAVAKTILG 18:  
Db 122 TLFKAGGSHSYALLAVSADAWKQTILG 148

## RESULT 6

C64721  
hypothetical protein b0011 - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: C64721; D56688; S28462  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: C64721  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-237 <BLAT>  
A:Cross-references: GB:AE000112; GB:U00396; NID:G1786192; PIDN:NAC73122.1; PID:G1786193;  
A:Experimental source: strain K-12, substrain MG1655  
R:James, R.; Dean, D.O.; Debbage, J.  
DNA Seq. 3, 327-332, 1993  
A:Title: Five open reading frames upstream of the dnaK gene of Escherichia coli.  
A:Reference number: A56688; MUID:94003405; PMID:8400364  
A:Accession: D56688  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-189, 'F', -91-237 <JAM>  
A:Cross-references: GB:X67700; NID:G41754; PIDN:CAA47934.1; PID:G41759

Query Match 20.7%; Score 265; DB 2; Length 237;  
Best Local Similarity 31.2%; Pred. No. 1.4e-12;  
Matches 77; Conservative 44; Mismatches 90; Indels 36; Gaps 6;  
QY 3 YKYDRDLFLKQLESSDLDLFEVLVFGKDGKGRK---HNEKLTSSIEYKR---HGDDYAKYAE 59  
Db 5 YLNDSDLDLFLQHCSEBQANFARLLTHNEKGRKRLSSVLNRLFKSMGHEPQHRNMQ 64  
QY 60 RIABELOYGNSPASFKEGVLKYLCDVCDKLVNKNKTTTTLIEQNMLSKILR 119  
Db 65 LIAGELQHFQGGDSIAKLRGKGLYRAILLDVSKRLKADKEMSTFEIEQQLLEQFLRN 124  
QY 120 SLEEMDDE-----EVKEMCDLSIKNTDNLNRLQALSATLTLFKMGGFKSYQLA 168  
Db 125 TWKKYDEEHKQBFHVAVDARVNELEBELLPLMKDLKLAGVS-----HLLSSQTL 174  
QY 169 VIVANAVAKTILGRGLSLAGNOVLRTLSFLTPGVWITGTWTAIDIAGPAYRVITPAC 228  
Db 175 RILRTHAAMSVLGHGL-LRG-----AGLGGPVGAAALNGVKA---VSGSAVRVITPAV 222  
QY 229 IVVATILR 235  
Db 223 LQIACLR 229

## RESULT 7

AC0503  
conserved hypothetical protein STY0010 [imported] - Salmonella enterica subsp. enterica  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: This species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AC0503  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
Th, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AC0503  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-237 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD01163.1; PID:G16501293; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY0010

Query Match 20.7%; Score 265; DB 2; Length 237;  
Best Local Similarity 32.6%; Pred. No. 1.4e-12;  
Matches 79; Conservative 40; Mismatches 101; Indels 22; Gaps 7;  
QY 1 MAYKYDRDLFLKQLESSDLDLFEVLVFGKDGKGRK---HNEKLTSSIEYKRHGDDY 54  
Db 3 VTYLHDEDLFLQHCSEBQANFARLLTHNEKGRKRLSSVLNRLFKAME--GHPEQH 59  
QY 55 AKYAEIAELOYGNSPASFKEGVLKYLCDVCDKLVNKNKTTTTLIEQNML 114  
Db 60 RRNWLQIAGEFQHYGGDSIAKLRGKGLYRAILLDVAKRLKADKSMSTFEIEQQLLE 119  
QY 115 KILERSLEEMDDDEVKEMCDLSIKNTDNLNRLQALSATLTLFK-MGGFKSYQLAVIVAN 173  
Db 120 HFLRHTWQKMDAAHKQBFLOAVDAKVSLEBELLPLMKDRSLKAGVSHLLSTQLTRI 179  
QY 174 AVAKTILGRGLSLAGNOVLRTLSFLTPGVWITGTWTAIDIAGPAYRVITPACIVVAT 233  
Db 180 HAAMSVLGHGL-LRG-----AGLGGPVGAAALNGVKA---MGSAYRVITPAVQIAC 227  
QY 234 LR 235  
Db 228 LR 229

## RESULT 8

C90630  
probable oxidoreductase ECs0012 [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C90630  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
Sagawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90630  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-237 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA033434.1; PID:G13359467; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECs0012

Query Match 20.5%; Score 262; DB 2; Length 237;  
Best Local Similarity 30.8%; Pred. No. 2.3e-12;  
Matches 76; Conservative 45; Mismatches 90; Indels 36; Gaps 6;  
QY 3 YKYDRDLFLKQLESSDLDLFEVLVFGKDGKGRK---HNEKLTSSIEYKR---HGDDYAKYAE 59  
Db 5 YLNDSDLDLFLQHCSEBQANFARLLTHNEKGRKRLSSVLNRLFKSMGHEPQHRNMQ 64  
QY 60 RIABELOYGNSPASFKEGVLKYLCDVCDKLVNKNKTTTTLIEQNMLSKILR 119  
Db 65 LIAGELQHFQGGDSIAKLRGKGLYRAILLDVSKRLKADKEMSTFEIEQQLLEQFLRN 124  
QY 120 SLEEMDDE-----EVKEMCDLSIKNTDNLNRLQALSATLTLFKMGGFKSYQLA 168  
Db 125 TWKKYDEEHKQBFHVAVDARVNELEBELLPLMKDLKLAGVS-----HLLSSQTL 174

RESULT 10  
F64718  
Hypothetical protein HP1590 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: F64718  
R:Author: R.Tomb, J.F.; White, C.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaki, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 519-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A6520; MIMD:97394467; PMID:9252185  
A:Accession: F64718  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-39 <TM>  
A:Cross-references: GB:AE000656; GB:AE000511; MIMD:g2314771; PIDN:AAD08629.1; PID:g2314771

Qy	201	GPVGWITGVWTAID	IAGPAYRV	TPACIVWATRL	KTQOANDKKS	LIQIESI	253
Db	915	G-----	GETFOSSLA	LALGLSEI	-----VQOQSS-	-----GISLESI	945

```

RESULT 12
AE0360
  hypothetical protein YPQ2963 [imported] - Yersinia pestis (strain CO92)
  C:Species: Yersinia pestis
  C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
  C:Accession: AE0360
  R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Brent
  deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dou
  il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.;
  Nature 413, 523-527, 2001
  A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
  A:Reference number: AE0001, MUID:21470413, PMID:11586360
  A:Accession: AE0360
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-282 <GR>
  A:Cross-references: KB:AL590842; PIDN:CAC92208.1; PID:q15980920; GSPDB:GN00175

```

RESULT 14  
C81380  
probable D-2-hydroxyacid dehydrogenase Cj0373 [imported] - Campylobacter jejuni (strain: C81380)  
C.Species: Campylobacter jejuni  
C.C-Date: 31-Mar-2000 #sequence\_revis:on 31-Mar-2000 #text\_change 03-Jun-2002  
C.C-Accession: C81380  
C.Reviewer: J. Wren, B. W. Mungali, K. Ketley, J. M. Churcher, C. Basham, D. Chillingworth, C. W. Quail, M. Rajandream, M. A. Rutherford, K. M. VanVliet, A. Whitehead, S. Barrett

Query Match	7.6%;	Score 97.5;	DB 2;	Length 728;
Best Local Similarity	25.4%;	Pred. No. 18;		
Matches	43;	Conservative 27;	Mismatches 56;	Indels 43; Gaps 8;
QY	10	EELKOLESSDLL---DLPELVFGKDGVRH---NEKUTSSIEY-----KRHGDDYAKY	57	
Dbb	352	KELIIVLTAKDLSSKGDLEEVIVYTKESLSRLVDLENIVKTDFLCVSAKMASDFYKGL	411	
QY	58	ARIEAEELQYGSNSFASPIKGEVLVYKEILCDVCDKLVNYKK-----TETTLTEQNM	112	
Dbb	412	ASK--ESLQSKQWQEF-----ENVLPNELYAGEKSIARLAYKKELHLEKKNILSEYEM	463	
QY	113	LSKILERSLEEMDDRE-----VKEMCDELS-----IKNTDN	143	
Dbb	464	QNRLIKENKQGVSEBNQKILLLEQNTLLKEAQDEISNIAKLKIDS	512	

Search completed: July 5, 2004, 03:13:15  
Job time : 14 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:08:23 ; Search time 12 seconds  
(without alignments)  
1097.812 Million cell updates/sec

Title: US-09-732-091-4

Perfect score: 1279

Sequence: 1 MAYKYDRDLFLKQLSSDL.....LRLKTOANGDKSLQIESI 253

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	1279	100.0	253	1	YF88_HELPY	026107 helicobacte
2	1270	99.3	253	1	YF88_HELPY	Q92124 helicobacte
3	637	45.8	209	1	YF87_HELPY	026106 helicobacte
4	630	45.3	209	1	YF87_HELPY	Q92125 helicobacte
5	265	20.7	237	1	YAAW_ECOLI	P75617 escherichia
6	262	20.5	237	1	YAAW_ECO57	P58316 escherichia
7	105	8.2	726	1	HS9A_BRARE	Q90474 brachydanio
8	97	7.6	916	1	SECA_TREPA	O83394 treponema p
9	96.5	7.5	373	1	BIOF_AQUAE	O66875 aquifex aeo
10	95.5	7.5	344	1	ABIC_LACLA	Q01457 lactococcus
11	95	7.4	517	1	YCD9_CAMJE	Q9pn86 campylobact
12	95	7.4	3214	1	BPA1_HUMAN	Q03001 homo sapien
13	94.5	7.4	1433	1	REST_CHICK	Q42184 gallus gall
14	93.5	7.3	802	1	PAC_BACME	Q60136 bacillus me
15	92.5	7.2	802	1	PAC_ARTVI	P31956 arthrobacte
16	92.5	7.2	892	1	RQCL_CABEL	P46364 caenorhabdi
17	91.5	7.2	399	1	RPSD_THEMA	P77994 thermotoga
18	91.5	7.2	472	1	GATE_METUA	Q57624 methanococc
19	91.5	7.2	719	1	HS9A_HORSE	Q9gkx7 equus cabal
20	91.5	7.2	1769	1	YXK9_YEAST	P42945 saccharomyc
21	91	7.1	284	1	TPM1_CIOIN	Q07068 clona intes
22	90	7.0	404	1	LA_BOVIN	Q927k0 chlamydia p
23	89.5	7.0	280	1	Y705_CHLNP	Q927k0 chlamydia p
24	89.5	7.0	633	1	GATE_SULSO	Q972h6 sulfolobus
25	89.5	7.0	869	1	AMPN_HAETN	P45274 haemophilus
26	89	7.0	386	1	NSG_HUMAN	Q9u116 homo sapien
27	88.5	6.9	732	1	HS9A_CRIGR	P46633 cricetus
28	88	6.9	371	1	RPSD_STRMG	O33662 streptococc
29	88	6.9	670	1	SSM4_SCHPO	O42667 schizosacch
30	87.5	6.8	728	1	HS9A_CHICK	P11521 gallus gall
31	87.5	6.8	731	1	HS9A_HUMAN	P07920 homo sapien
32	87.5	6.8	732	1	HS9A_MOUSE	P07901 mus musculu
33	87.5	6.8	732	1	HS9A_PIG	O02705 sus scrofa

34 87 6.8 607 1 HTPG\_FUSNN Q8rgh4 fusobacteri  
35 86.5 6.8 619 1 PRIA\_HELPY Q9zke4 helicobacte  
36 86.5 6.8 686 1 HMCT\_HELPY Q9z153 helicobacte  
37 86.5 6.8 1076 1 CARB\_ARCFU Q28994 archaeoglob  
38 86.5 6.8 1102 1 PRIA\_STRCO Q9kxr6 streptomyc  
39 86 6.7 619 1 PRIA\_HELPY Q25149 helicobacte  
40 86 6.7 633 1 YN40\_ARCFU Q27944 archaeoglob  
41 86 6.7 1164 1 KELL\_YEAST P38853 saccharomyc  
42 86 6.7 1169 1 EX5B\_BORBU O51578 borrelia bu  
43 85.5 6.7 411 1 HFLK\_BUCAP Q8k914 buchnera ap  
44 85.5 6.7 1085 1 AMPI\_PLAFO Q96935 plasmodium  
45 85 6.6 894 1 SEC3\_HUMAN Q9nv70 homo sapien

## ALIGNMENTS

### RESULT 1

YF88\_HELPY  
ID YF88\_HELPY STANDARD; PRT; 253 AA.  
AC O26107;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein HP1588.  
GN HP1588.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=26695 / ATCC 700392;  
RX MEDLINE=97394467; Pubmed=9252185;  
RA Tomb J.-F., White O., Kervage A.R., Clayton R.A., Sutton G.G.,  
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.V.,  
RA Venter J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
pylori.";  
RL Nature 388:539-547(1997).  
CC -!- SIMILARITY: Belongs to the UPF0174 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE000656; RAD08627.1; -;  
DR PIR; D64718; D64718.  
DR TIGR; HP1588; -;  
DR InterPro; IPR005367; UPF0174.  
DR Pfam; PF03667; UPF0174; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 253 AA; 28417 MW; 00E15A38B1A2036A CRC64;

Query Match 100.0%; Score 1279; DB 1; Length 253;  
Best Local Similarity 100.0%; Pred. No. 9.1e-88;  
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYKYDRDLFLKQLSSDLFLFLVFGKDGKRNHEKLTSSIEYKRGHGDYAKYAE 60  
Db 1 MAYKYDRDLFLKQLSSDLFLFLVFGKDGKRNHEKLTSSIEYKRGHGDYAKYAE 60  
QY 61 IAEELQYVGSNSFASFKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

```

Db 61 IAELOQYSGNSFASFKGEGVLYKEILCDVCDKLKVNKKTTTLLIEQNMLSKILERS 120
QY 121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSATLTLPKMGFKSYQLAVIVANAVAKTIL 180
Db 121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSATLTLPKMGFKSYQLAVIVANAVAKTIL 180
QY 181 GRGLSLAGNOVLTRTSLFLGPGVGIITGVWTAIDIAAGPAYRVITIPACIVVATRLKTKQ 240
Db 181 GRGLSLAGNOVLTRTSLFLGPGVGIITGVWTAIDIAAGPAYRVITIPACIVVATRLKTKQ 240
QY 241 ANGDKKSLOLES 253
Db 241 ANGDKKSLOLES 253

RESULT 2
YF88_HELPJ
ID YF88_HELPJ STANDARD; PRT; 253 AA.
AC Q3ZJ24;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein JHP1494.
GN JHP1494.
OS Helicobacter pylori J93 (Campylobacter pylori J93).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120537; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown B.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Malls S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
RN [2]
CC -1- SIMILARITY: Belongs to the UPF0174 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; A5001571; A5007073.1;
CC PIR; B71800; B71800.
CC InterPro; IPR005367; UPF0174.
CC Pfam; PF03667; UPF0174; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 253 AA; 28475 MW; 127158B2B1A20336A CRC64;

Query Match 99.3%; Score 1270; DB 1; Length 253;
Best Local Similarity 99.2%; Pred. No. 4.2e-87;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAYKYDRDLFEFLKQLLESSLLDLFEVLVFGKDGKEKHNEKLTSSIEYKSHGDDYAKYAE 60
Db 1 MAYKYDRDLFEFLKQLLESSLLDLFEVLVFGKDGKEKHNEKLTSSIEYKSHGDDYAKYAE 60
QY 61 IAELOQYSGNSFASFKGEGVLYKEILCDVCDKLKVNKKTTTLLIEQNMLSKILERS 120
Db 61 IAELOQYSGNSFASFKGEGVLYKEILCDVCDKLKVNKKTTTLLIEQNMLSKILERS 120
QY 121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSATLTLPKMGFKSYQLAVIVANAVAKTIL 180
Db 121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSATLTLPKMGFKSYQLAVIVANAVAKTIL 180

```

```

QY 181 GRGLSLAGNOVLTRTSLFLGPGVGIITGVWTAIDIAAGPAYRVITIPACIVVATRLKTKQ 240
Db 181 GRGLSLAGNOVLTRTSLFLGPGVGIITGVWTAIDIAAGPAYRVITIPACIVVATRLKTKQ 240
QY 241 ANGDKKSLOLES 253
Db 241 ANGDKKSLOLES 253

RESULT 3
YF87_HELPY
ID YF87_HELPY STANDARD; PRT; 209 AA.
AC O26106;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein HP1587.
GN HP1587.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX XEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
RN [2]
CC CONCEPTUAL TRANSLATION.
CC Bairoch A.;
CC Unpublished observations (OCT-2001).
CC -1- SIMILARITY: Belongs to the UPF0174 family.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 160.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; A5000656; A500626.1; ALT_FRAME.
CC TIGR; HP1587;
CC InterPro; IPR005367; UPF0174.
CC Pfam; PF03667; UPF0174; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 209 AA; 23069 MW; F98D3FB8F3F62323 CRC64;

Query Match 49.8%; Score 637; DB 1; Length 209;
Best Local Similarity 60.3%; Pred. No. 2.3e-40;
Matches 123; Conservative 33; Mismatches 44; Indels 4; Gaps 2;

QY 38 NEKLSSIEYKSHGDDYAKYAEIABELOYGNSPASFKGEGVLYKEILCDVCDKLK 97
Db 2 NEKLSSIEYKSHGDDYAKYAEIABELOYGNSPASFKGEGVLYKEILCDVCDKLK 97
QY 98 NYNKKTTTLLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSATLT 154
Db 62 NYNKSATSLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSATLT 121

```



```
YAAW EC057
ID YAAW_EC057 STANDARD; PRT; 237 AA.
AC P58316;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein YAAW.
GN YAAW OR 20011 OR RCS0012.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22 (2001).
RN [3]
RP SEQUENCE FROM N.A.
CC -!- SIMILARITY: Belongs to the UPF0174 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AE005178; AG54311.1; -
CC EMBL; AF002550; BAB3434.1; -
CC PIR; C85481; C85481.
CC InterPro; IPR005367; UPF0174.
CC Pfam; PF03667; UPF0174; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 237 AA; 26681 MW; A25482B2E116759 CRC64;

Query Match 20.5%; Score 262; DB 1; Length 237;
Best Local Similarity 30.8%; Pred. No. 1, 5e-12;
Matches 76; Conservative 45; Mismatches 90; Indels 36; Gaps 6;

QY 3 YKDYDLFLKQLESDDLDFVLVFGKGRHNEKLTSSIEYR-----HGGDYAKYAE 59
DB 5 YLNDLDFLQCBSBEQANFARLLTHNEKGRLLSSVLNRLNLFKSMGHPHQHRRNQ 64
QY 60 RIAEELQYSGNSFASFKIGEVLYKEILCDVCDKLVKNYKTTTLIEQNMLSKILER 119
DB 65 LIAGELOHFGGSIANKLGHKLYRAILLDSVSKLKKADKEMSTFEIQQLLEQFLRN 124
QY 120 SLEEMDDE-----EVKEMCDLSIKNTDNLNROALSAAATLTFPMGCFKSYQLA 168
DB 125 TWKKMDEEHKQEFLLHVAQVAVNELEELLPLMKDKLLAGVS-----HLLSSQLT 174
QY 169 VIVANAVAKTIILGRGLSLAGNQLVRLTSLFLGPGVGIITGWTAIDAGPAYRVTIPAC 228
DB 175 RILRTHAAMSVLGHGL-LRG-----AGLGGPVGAALNGVKA---VSSSSYRVTIPAV 222

YAAW EC057
ID YAAW_EC057 STANDARD; PRT; 726 AA.
AC Q90474;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heat shock protein HSP 90-alpha.
GN HSP90A OR HSP90.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC Lele Z., Hadfi S., Sass J.B., Krone P.H.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 36-136 FROM N.A.
RX MEDLINE=95071389; PubMed=7980538;
RA Krone P.H., Sass J.B.;
RT "HSP 90 alpha and HSP 90 beta genes are present in the zebrafish and
RT are differentially regulated in developing embryos.";
RL Biochem. Biophys. Res. Commun. 204:746-752 (1994).
CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the heat shock protein 90 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AF068773; AAC21567.1; -
CC EMBL; L35586; AAA97518.1; -
CC PIR; JC2343; JC2343.
CC HSSP; P07900; 1BYQ.
CC ZFIN; ZDB-GENE-990415-94; hsp90a.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR001404; Hsp90.
CC Pfam; PF02518; HATPase_c; 1.
CC Pfam; PF00183; HSP90; 1.
CC PRINTS; PR00775; HEATSHOCK90.
CC SMART; SM00387; HATPase_c; 1.
CC PROSITE; PS00298; HSP90_1.
KW Chaperone; ATP-binding; Heat shock.
SQ SEQUENCE 726 AA; 83561 MW; F7DEB8EF1FEC9CEC CRC64;

Query Match 8.2%; Score 105; DB 1; Length 726;
Best Local Similarity 21.4%; Pred. No. 2, 5;
Matches 61; Conservative 41; Mismatches 75; Indels 108; Gaps 13;

QY 21 LDLEVLVFGKGRHNEKLTSSIEYKRGDDYAKYAEIAYELQYSGNS----- 72
DB 415 LDLELAEADKQNYKYEQFSKNIKGLIHED--SQNRKKLSDLLRYTTSASGDEMVS 472
QY 73 -FAS-----FLKGE-----GVLYK-ELCDVCDKLVKNYK 101
DB 473 DVVSRMKOTQKHYYITGETKQVQVANSFAVEKLRKAGLEVITYMIEPIDYCVQQLKEY 532
QY 102 KTETTLIEQNMLSKILERSLEEMDDEEVKEMCDLSIKNTDNLNROALSAAATLTFPMG 161
```

Db 533 K-----NLVS-VTKGLELPEDREKKQDELKAK-YENLCK-----567  
 QY 162 FKSQYLAIVANAVAKTILGRGUS--LAGNQVLTSLTSLTGPVGHIIIT-----208  
 Db 566 -----IMKHIDKKIEKVTVSNRLVSSECCIVTSTYGTANMERIMKSQLR 614  
 QY 209 -----GVWTF-----IDIAGPAYRTIPACIVAVATRLKLTQOANGK 245  
 Db 615 DNSTGWTAKKHLEIN-----PAHPIVETLREKAEKNDK 651

RESULT 8  
 SSCA TREPA  
 ID SECA TREPA STANDARD; PRT; 916 AA.  
 AC O83394.  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Preprotein translocase secA subunit.  
 GN SECA OR TP0379.

OS Treponema pallidum.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nichols;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 Sodergren E., Harcham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 Khalak H., Richardson P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 McDonald L., Artlach P., Howell J.K., Chidambaram M., Uterback T.,  
 Hatch B., Horst K., Roberts K., Sardusky M., Weidman J., Smith H.O.,  
 Venter J.C.;  
 RA "Complete genome sequence of Treponema pallidum, the syphilis  
 spirochete.";  
 RL Science 281:375-388(1998).  
 CC -!- FUNCTION: Involved in protein export. Interacts with the secY/secE  
 subunits. SecA has a central role in coupling the hydrolysis of  
 ATP to the transfer of pre-secretory periplasmic and outer  
 membrane proteins across the membrane (By similarity).  
 CC -!- SUBUNIT: Part of the prokaryotic protein translocation apparatus  
 which comprise secA, secB, secE, secF, secG and secY (By  
 similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic side of plasma membrane (By  
 similarity).  
 CC -!- SIMILARITY: Belongs to the secA family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EXBL; AE001217; AAC65365.1; --  
 CC PIR; E71330; E71330.  
 CC TIGR; TP0379.  
 CC InterPro; IPR004027; SEC\_C\_motif.  
 CC InterPro; IPR000185; SecA.  
 CC Pfam; PF02810; SEC-C; 1.  
 CC Pfam; PF01043; SecA\_protein; 1.  
 CC PRINTS; PR00906; SECA.  
 CC TIGRFAMs; TIGR00963; secA; 1.  
 CC PROSITE; PS01312; SEC\_A; 1.  
 CC Protein transport; ATP-binding; Vembrane; Translocation; Transport;  
 CC Complete proteome.  
 FT NP BIND 103 110 ATP (POTENTIAL).  
 SQ SEQUENCE 916 AA; 103779 MW; EA5561F6EE7C65AE CRC64;

Query Match 7.6%; Score 97; DB 1; Length 916;  
 Best Local Similarity 21.8%; Pred. No. 13;

Matches 52; Conservative 44; Mismatches 67; Indels 76; Gaps 13;  
 QY 3 YKTRDLEFLKQLESDLLDLFEVLVFGKDE-----KRNHEKLTSSIEYK 48  
 Db 311 EKY---IHVPTQALRAHLVLRADYVVKDQVQIVDEFTGRILLEGRRYSGLHQAIEAK 367  
 QY 49 RHGDDYAKVAERIAEELQYSGNSPASFIKGEVLVYKEI--LQDVCKLKVNYNK--KTE 104  
 Db 368 EH-----IRIAQRNRMTATITFONFFR---MYKLSGMTGTADTEALELNKIYKLE 415  
 QY 105 TTLIEQNM--LSKILERSLEMDDEEV--KEMCDE-----LSIKNTLNLRQ 147  
 Db 415 VVVLPTNLPVARVDEHDVVVLSSEKWSAICDEIKAEHTRGQPVVLTISSEKSEK-- 473  
 QY 149 ALSAATLTIFKMGPKSYQL-----AVIVANA-----VAKTILGRG--LSLAGN 189  
 Db 474 -----ALLTRGVKGVHELVNAKNHAREALIIAEAGAKSGVTIATNAGSGTDIKLGN 525

## RESULT 9

BIOP AQUAE  
 ID BIOP AQUAE STANDARD; PRT; 373 AA.  
 AC O66875;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 8-amino-7-oxononanoate synthase (EC 2.3.1.47) (AONS); (8-amino-7-  
 keto-pelargonic acid synthetase)  
 GN (7-KAP synthetase) (L-alanine--pimelyl CoA ligase).  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
 Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RA "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.";  
 RT Nature 392:353-358(1998).  
 RL Nature 392:353-358(1998).  
 CC -!- CATALYTIC ACTIVITY: 6-carboxyhexanoyl-CoA + L-alanine = 8-amino-7-  
 oxononanoate + CoA + CO(2).  
 CC -!- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -!- PATHWAY: Biotin biosynthesis; first step.  
 CC -!- SIMILARITY: Belongs to class-II of pyridoxal-phosphate-dependent  
 aminotransferases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE000699; AAC06836.1; --  
 CC PIR; G70355; G70355.  
 CC HSSP; P12998; 1BS0.  
 CC InterPro; IPR003408; Ala\_synthase.  
 CC InterPro; IPR004839; Aminotrans\_I/II.  
 CC InterPro; IPR001917; Aminotrans\_II.  
 CC InterPro; IPR004723; Biof.  
 CC Pfam; PF02490; ALA synthase; 1.  
 CC Pfam; PF00155; aminotran\_1\_2; 1.  
 CC TIGRFAMs; TIGR00858; bioF; 1.  
 CC PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
 CC Biotin biosynthesis; Transferase; Pyridoxal phosphate;  
 CC Complete proteome.  
 FT BINDING 223 223 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 373 AA; 42532 MW; 536B34ASD5F84401 CRC64;



Query Match 7.5%; Score 96.5; DB 1; Length 373;  
 Best Local Similarity 22.6%; Pred. No. 4.7;  
 Matches 64; Conservative 40; Mismatches 102; Indels 77; Gaps 14;

QY 9 LEFLKQLLESSDLDFEVLVFGDKGKRHNKLTSSIEYKRGH-----DYAKYA 58  
 DB 23 LEGVKQFCSNDYL-----GLRKKPEVVEESIRVLKRGAGLGSQALVSGYTKHH 71  
 QY 59 ERIAEEL-QYVGSNSFASFIKG-----EGVLVKEILCDVCCKLKVNNYKTE 104  
 DB 72 RELEEKLAEPKGTSCVLFSGFLANWGVTIPALVEGDL--VLSD-----ELNH----- 118  
 QY 105 TTLEQNMLSKILERSLEEMDDDEVKMCDELSTKNTDNLNRQALSATLTLFKMGCG-- 161  
 DB 119 ASIIGVRLSKAKRQKFKKKDYBELSEFLK-----KNRKKFRVLI--ITDTVFSMDGDA 172  
 QY 162 -----FKSYQLAVIVANAVAKTILGR-GLSLAGNQ-----VLNRLTSLFTGPGWMI 207  
 DB 173 DLXELTQICEYDCMLYIDEAHTTGTGKGLDYFGIEHKEVYIYVNGTLKSKALGSYGAFV 232  
 QY 208 TGWTAIDIAGRAYRVTIPACIVATVRLKLTQOAGKSKSLQI 250  
 DB 233 CGTKLLID-----YLVNKARSLFST-SLPPSVCAKAKAI 269

RESULT 10  
 ABIC\_LACIA STANDARD; PRT; 344 AA.  
 AC Q01457;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE Abortive phage resistance protein abic.  
 GN ABIC OR PRF.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OG Plasmid pTN20.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1360;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ME2;  
 RC MEDLINE=93054365; PubMed=1429469;  
 RA Durmaz E., Higgins D.L., Klaenhammer T.R.;  
 RT "Molecular characterization of a second abortive phage resistance gene present in Lactococcus lactis subsp. lactis ME2".  
 RL J. Bacteriol. 174:7463-7469 (1992).  
 CC - FUNCTION: PROVIDES RESISTANCE TO BACTERIOPHAGE BY ABORTIVE INFECTION.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M95956; AAA53569.1; --  
 CC PIR; A47025; A47025.  
 CC Plasmid.  
 CC  
 CC SEQUENCE 344 AA; 40128 MW; AA6E41E71B4A003C CRC64;  
 DR PIR; A47025; A47025.  
 DR Plasmid.  
 DR  
 DR SEQUENCE 344 AA; 40128 MW; AA6E41E71B4A003C CRC64;  
 Query Match 7.5%; Score 95.5; DB 1; Length 344;  
 Best Local Similarity 23.8%; Pred. No. 5.1;  
 Matches 50; Conservative 36; Mismatches 73; Indels 51; Gaps 9;

QY 7 RLEFLKQLLESSD-----LLDFEVLVFGDKGKRHNKLTSSIEYKX-----HGDD 53  
 DB 92 KNQEESKLLNNSANREFYSLDLFX-----KEQNKSETIKALSFL-YKRAINDKHGNS 144  
 QY 54 YAK-----YAEIAEELQYVGSNSFASFIKGGLVYKEILCDVCCKLKVNNYKX 102

Query Match 7.4%; Score 95; DB 1; Length 517;  
 Best Local Similarity 21.2%; Pred. No. 9.1;  
 Matches 54; Conservative 43; Mismatches 98; Indels 60; Gaps 10;

QY 2 AYKYDRDLEFLKQLLESSDLDFEVLVFGDKGKRHNKLTSSIEYKRGHDDYAKYAER 60  
 DB 43 AIEYAEEL-IKDKANSILNAELSV-----KKYKETHKKFQKDFNKFDDLSKKEOK 94

Query Match 7.4%; Score 95; DB 1; Length 517;  
 Best Local Similarity 21.2%; Pred. No. 9.1;  
 Matches 54; Conservative 43; Mismatches 98; Indels 60; Gaps 10;

QY 2 AYKYDRDLEFLKQLLESSDLDFEVLVFGDKGKRHNKLTSSIEYKRGHDDYAKYAER 60  
 DB 43 AIEYAEEL-IKDKANSILNAELSV-----KKYKETHKKFQKDFNKFDDLSKKEOK 94

```

CY 61 IABLEQYVGSNSFASFIKGGVLYKELC-----DVCCKLVNNYKKTFTTL-- 107
DB 95 LQEECKEKED-----KEYLCKSQKHQNLQSDV-DKLNKYQEKLDVVLKI 140
CY 108 -----IFONMLSKTLERSLEEMDEEVKEMCDELSIKNTKLNROALSAATLTLFRMG 161
DB 141 LEHSTGLTQNAKEILKXVNSREQIAHIVKYEEAKNEAKKANFIQAQTSFRAG 200
CY 162 FKSQVLAVIVANAVAKTILGRGLSAGNQVLTRTLSPITGPVGVWITGTWTAIDIAGPAY 221
DB 201 EFAAERLNVINIKNDLKGRIIGREGRNV--KTLG-----MVLGVDDIID----- 244
CY 222 RVTPACVWATLRL 236
DB 245 --DTPGAIIVSCFNL 257

RESULT 12
BPAL HUMAN
ID _BPAL_HUMAN STANDARD; PRT; 3214 AA.
AC Q03001; Q12825; Q13266; Q13775; Q96J76; Q96Q75; Q9UGD7;
AC Q9UGD8;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1 isoforms 1/2/3/4/5/8 (230 kDa bullous
DE pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein) (Dystonia
DE Musculorum protein) (Fragment).
GN BPAG1 OR DMH OR DT OR KIAA0728.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
R2 SEQUENCE FROM N.A. (ISOFORM 3).
R3 TISSUE=Keratinocytes;
RX MEDLINE=92011493; PubMed=1717441;
RA Sawamura D., Li X., Chu M.-L., Uitto J.;
RT "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences
RT deduced from cloned cDNAs predict biologically important peptide
RT segments and protein domains.";
RL J. Biol. Chem. 266:17734-17790(1991).
RN [2]
R2 SEQUENCE FROM N.A. (ISOFORM 3).
R3 TISSUE=Keratinocytes;
RX MEDLINE=93346806; PubMed=8345227;
RA Elgart G.W., Stanley J.R.;
RT "Cloning of the 5' mRNA for the 230-kD bullous pemphigoid antigen by
RT rapid amplification of cDNA ends.";
RL J. Invest. Dermatol. 101:244-246(1993).
RN [3]
R2 SEQUENCE OF 1-645 FROM N.A. (ISOFORMS 1 AND 2).
R3 TISSUE=Fetal brain, and Retina;
RX MEDLINE=96213394; PubMed=8575775;
RA Brown A., Dalpe G., Mathieu M., Kothary R.;
RT "Cloning and characterization of the neural isoforms of human
RT dystonin.";
RL Genomics 29:777-780(1995).
RN [4]
R2 SEQUENCE OF 321-3214 FROM N.A. (ISOFORMS 1 AND 4).
RA Laird G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
R2 SEQUENCE OF 392-492 FROM N.A.
R3 TISSUE=Pineal gland;
RA Geerts D., Sonnenberg A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
R2 SEQUENCE OF 1082-3214 FROM N.A. (ISOFORM 3).
R3 TISSUE=Keratinocytes;
RX MEDLINE=91286285; PubMed=1712022;
RA Tanaka T., Parry D.A.D., Klaus-Kovtun V., Steinert P.M.,

```

```

RA Stanley J.R.;
RT "Comparison of molecularly cloned bullous pemphigoid antigen to
RT desmoplakin I confirms that they define a new family of cell adhesion
RT junction plaque proteins.";
RL J. Biol. Chem. 266:12555-12559(1991).
RN [7]
R2 SEQUENCE OF 2160-2767 FROM N.A.
R3 MEDLINE=91216368; PubMed=2090522;
RA Owaribe K., Kartenbeck J., Stumpp S., Magin T.M., Krieg T.,
RA Diaz L.A., Franke W.W.;
RT "The hemidesmosomal plaque. I. Characterization of a major
RT constituent protein as a differentiation marker for certain forms of
RT epithelia.";
RL Differentiation 45:207-220(1990).
RN [8]
R2 SEQUENCE OF 2287-3214 FROM N.A. (ISOFORM 5).
R3 TISSUE=Keratinocytes;
RX MEDLINE=89067122; PubMed=2461961;
RA Stanley J.R., Tanaka T., Mueller S., Klaus-Kovtun V., Roop D.;
RT "Isolation of complementary DNA for bullous pemphigoid antigen by use
RT of patients' autoantibodies.";
RL J. Clin. Invest. 82:1864-1870(1988).
RN [9]
R2 SEQUENCE OF 2462-3214 FROM N.A. (ISOFORM 8).
R3 MEDLINE=94280413; PubMed=8010969;
RA Hopkinson S.B., Jones J.C.;
RT "Identification of a second protein product of the gene encoding a
RT human epidermal autoantigen.";
RL Biochem. J. 300:851-857(1994).
RN [10]
R2 DOMAINS
R3 MEDLINE=96199235; PubMed=8621649;
RA Tang H.-Y., Chaffotte A.-F., Thacher S.M.;
RT "Structural analysis of the predicted coiled-coil rod domain of the
RT cytoplasmic bullous pemphigoid antigen (BPAG1). Empirical
RT localization of the N-terminal globular domain-rod boundary.";
RL J. Biol. Chem. 271:9716-9722(1996).
RN [11]
R2 SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE SPLICING.
R3 TISSUE=Keratinocytes;
RX MEDLINE=21839111; PubMed=11751855;
RA Okumura M., Yamakawa H., Ohara O., Owaribe K.;
RT "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1)
RT including the domain structure closely related to MACP (microtubule
RT actin cross-linking factor).";
RL J. Biol. Chem. 277:6682-6687(2002).
RN [12]
R2 FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
R3 intermediate filaments to the inner plaque of hemidesmosomes. The
R3 proteins may self-aggregate to form filaments or a two-dimensional
R3 mesh.
R3 SUBUNIT: Homodimer.
R3 SUBCELLULAR LOCATION: Cytoplasmic.
R3 ALTERNATIVE PRODUCTS:
R3 Event=Alternative splicing; Named isoforms=10;
R3 Comment=Isoforms 1, 2, 5 and 8 are or may be fragments;
R3 Name=1;
R3 IsoId=Q03001-1; Sequence=Displayed;
R3 Name=2;
R3 IsoId=Q03001-2; Sequence=VSP_005053, VSP_005055;
R3 Name=3; Synonyms=1e;
R3 IsoId=Q03001-3; Sequence=VSP_005054, VSP_005056, VSP_005057,
R3 VSP_005058, VSP_005059, VSP_005060,
R3 VSP_005061;
R3 Name=4;
R3 IsoId=Q03001-4; Sequence=VSP_005054, VSP_005056;
R3 Name=5;
R3 IsoId=Q03001-5; Sequence=VSP_005064, VSP_005065;
R3 Name=6; Synonyms=EA;
R3 IsoId=Q04833-2; Sequence=External;
R3 Name=7; Synonyms=EB;
R3 IsoId=Q08WXX8-1; Sequence=External;
R3 Name=8;
R3 IsoId=Q03001-6; Sequence=VSP_005062, VSP_005063;

```



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC  EMBL; AF014012; AAC60344.1; -
CC  EMBL; AF020764; AAC60345.1; -
CC  EMBL; AF045650; AAC03547.1; -
CC  EMBL; AF045651; AAC03548.1; -
CC  InterPro; IPR000938; CAP-Gly.
CC  InterPro; IPR001878; Znf.CCHC.
CC  Pfam; PF01302; CAP_GLY; 2.
CC  SMART; SM00343; Znf.C2HC; 1.
CC  PROSITE; PS00845; CAP_GLY_1; 2.
CC  PROSITE; PS02445; CAP_GLY_2; 2.
KW  Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
FT  DOMAIN 79 121 CAP-GLY 1.
FT  DOMAIN 144 207 SER-RICH.
FT  DOMAIN 235 277 CAP-GLY 2.
FT  DOMAIN 305 332 SER-RICH.
FT  DOMAIN 351 1353 COILED COIL (POTENTIAL).
FT  DOMAIN 1414 1427 CCHC-BOX.
FT  VARSPLIC 458 492 Missing (in isoform 2).
FT  VARSPLIC 458 492 /FTID=VSP 000761.
FT  VARSPLIC 458 492 /FTID=VSP 000761.
FT  VARSPLIC 803 803 S -> GSSKVS (in isoform 3).
FT  VARSPLIC 458 458 T -> RKQIREDPENT (in isoform 4).
FT  CONFLICT 309 309 K -> R (IN REF. 2; AAC03547).
FT  CONFLICT 440 440 E -> V (IN REF. 2; AAC03548).
SQ  SEQUENCE 1433 AA; 161026 MW; 5631CE8683498823 CRC64;

Query Match
Best Local Similarity 7.4%; Score 94.5; DB 1; Length 1433;
Matches 52; Conservative 36; Mismatches 74; Indels 51; Gaps 10;

QY 1 MAYKDRDLFLKQLSSLDLDELVEVFGKDGKRNHKLTSSTYEYKGGCDYAKYAE 60
DB 1020 MITHKADIKGFKQ-----NLLDAEAL---KAAQKNDLELTQAEELKQAEQ-AKADKR 1071

QY 61 IAEELQYGNNSFASFIKGGVLYK---EILCDVCDKLKVNYNKKTETTLTEQNMLSKIL 117
DB 1072 ABEVLQ-----TMEKVTKEKDAIHQEKIETLASLNGRQTKELQNELDMLKQNNLNKEE 1126

QY 118 E-----RSLEEMDEEVKEMCELSIKNTDNLNRQALSAATILTFKMGGFKSYQLAVI-VA 172
DB 1127 ELTKSKELLNLNKKVLELKKPE-----EALKLAAQ-----KSQLAALQEE 1169

QY 173 NAVAKTILGRG-----LSLAGNQVL 192
DB 1170 NVKLABELGRSDRBTWTSQKLEERSVLYNNQLL 1202

RESULT 14
PAC_BACME
ID PAC_BACME STANDARD; PRT; 802 AA.
AC Q60136; Q9S463;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE (Penicillin G amidohydrolase).
GN PAC OR PGA.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14945;
RA Kang J.H., Kim S.J., Park Y.C., Huang Y., Yoo O.J., Kim Y.C.;
RT "Nucleotide sequence of the penicillin G acylase gene from
Bacillus megaterium and characteristics of the enzyme."

```

```

RL Misainmurhag Hoiiji 32:215-221(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14945;
RX MEDLINE=95180705; PubMed=7875576;
RT Martin L.M., Prieto A.M., Cortes E., Garcia J.L.;
RA "Cloning and sequencing of the pac gene encoding the penicillin G
acylase of Bacillus megaterium ATCC 14945.";
RT FEMS Microbiol. Lett. 125:287-292(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CA4098;
RX PubMed=12114980;
RT Yang S., Huang Y.H., Huang X.D., Li S.Y., Yuan Z.Y.;
RA "High expression of penicillin G acylase gene from Bacillus megaterium
in Bacillus subtilis.";
RT Acta Biochim. Biophys. Sin. 31:601-603(1999).
RC -!- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-
aminopenicillanate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC -!- SUBUNIT: Heterodimer of an alpha chain and a beta chain (By
similarity).
CC -!- SUBCELLULAR LOCATION: Extracellular (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S45.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```

CC  EMBL; U07682; AAB41343.1; -
CC  EMBL; Z37542; CA85774.1; -
CC  EMBL; AF161313; AAD45609.1; -
CC  PIR; S49252; S49252.
CC  HSSP; P06875; 1AJQ.
CC  MEROPS; S45.001; -.
CC  InterPro; IPR002692; Peptidase_S45.
CC  Pfam; PF01804; Penicil_amidase; 1.
CC  Hydrolase; Antibiotic resistance; Zymogen; Calcium-binding; Signal.
CC  SIGNAL 1 24 POTENTIAL.
CC  CHAIN 25 802 PENICILLIN G ACYLASE ZYMOGEN.
CC  CHAIN 25 234 PENICILLIN G ACYLASE ALPHA SUBUNIT.
CC  PROPEP 235 265 SPACER PEPTIDE.
CC  CHAIN 266 802 PENICILLIN G ACYLASE BETA SUBUNIT.
CC  ACT_SITE 266 266 BY SIMILARITY.
CC  METAL 177 177 CALCIUM (POTENTIAL).
CC  METAL 341 341 CALCIUM (POTENTIAL).
CC  VARIANT 3 3 T -> M (IN STRAIN CA4098).
CC  VARIANT 224 224 D -> E (IN STRAIN CA4098).
CC  VARIANT 232 232 I -> K (IN STRAIN CA4098).
CC  VARIANT 254 254 T -> S (IN STRAIN CA4098).
CC  VARIANT 349 349 A -> T (IN STRAIN CA4098).
CC  VARIANT 470 470 S -> N (IN STRAIN CA4098).
CC  VARIANT 524 524 F -> Y (IN STRAIN CA4098).
CC  VARIANT 569 569 Q -> P (IN STRAIN CA4098).
CC  VARIANT 586 586 I -> A (IN STRAIN CA4098).
CC  VARIANT 657 657 N -> S (IN STRAIN CA4098).
CC  VARIANT 740 740 T -> K (IN STRAIN CA4098).
CC  VARIANT 789 791 NKA -> YKS (IN STRAIN CA4098).
SQ SEQUENCE 802 AA; 91987 MW; 877CA0564E50DFBD CRC64;

```

Query Match  
Best Local Similarity 7.3%; Score 93.5; DB 1; Length 802;  
Matches 62; Conservative 39; Mismatches 107; Indels 77; Gaps 10;

```

QY 4 KYDRDLFLKQLSSLDLDFEV-LVFGKDG-----EKRNHKLTSSTYEYKGGDDYA 55
DB 33 KVVDRNFGVPHLVAKNKKDLYEAYGVYMAKDRLFQLEMPFRGNEGTVSEI-----FGEDYL 88
QY 56 -----KYAERIAELQYGNNSFASFIKGGVLYKEILCDVCDKLKVNYN-- 100

```

```
Db 89 SKDEQSRDGYSNKEIKKMDGLDROPKELLIAFAEGISRYVNEALKDPDDKLSKEFHEY 148
QY 101 -----KKTETTLIEQNMLS-----KILERSLEEMDDBEVKEMCDELSTKN- 140
Db 149 QFLPQKWTSTDVVRVYVMSVNTYPMFNKHQELAKNAEILAKLEHEYGTEVSRKXCFDDLVWKND 208
QY 141 -----TDNLNRQALSAATLTLFKMGFK-----SYQLAVIVANAVAK 177
Db 209 PSAPTIVSEGGPKRDSQSLQILSSAVIKASEKVKGRNENFVQTSSEELGLPLKTSNA 268
QY 178 TILGRGLSLAGNOVLRTLSFLTGP-VGWIITGVWTAIDIAIPAY 221
Db 269 AIWGEKSEKATGNALL-----FSGPQGVFVAPGFLYEVLHAPGF 307

RESULT 15
PAC_ARTVI
ID PAC_ARTVI STANDARD; PRT; 802 AA.
AC P31956;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE (Penicillin G amidohydrolase).
GN PAC OR PA.
OS Arthrobacter viscosus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococccineae; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=1673;
RW [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15294;
RX MEDLINE=94259306; PubMed=8200542;
RA Konstantinovic M., Marjanovic N., Ljubijankic G., Glisin V.;
RT "The penicillin amidase of Arthrobacter viscosus (ATCC 15294).";
RL Gene 143:79-83(1994).
CC -!- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-
CC aminopenicillanate.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC -!- SUBUNIT: Heterodimer of an alpha chain and a beta chain (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Extracellular (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S45.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: L04471; AAA22077.1; -.
CC PIR: I39665; I39665.
CC HSP: P08375; LAJQ.
CC MEROPS: S45.001; -.
CC InterPro: IPR002692; Peptidase S45.
CC Pfam: PF01804; Penicil amidase; 1.
CC Hydrolase; Antibiotic Resistance; 1.
CC SIGNAL 1 24
CC CHAIN 25 802
CC CHAIN 25 234
CC PROPEP 235 265
CC CHAIN 266 802
CC ACT SITE 266 266
CC METAL 177 177
CC METAL 341 341
CC SEQUENCE 802 AA; 92113 MW; 9863E58C526C85D7 CRC64;
```

Query Match 7.2%; Score 92.5; DB 1; Length 802;  
Best Local Similarity 21.8%; Pred. No. 24;  
Matches 62; Conservative 39; Mismatches 107; Indels 77; Gaps 10;

```
QY 4 KYDRDLFLKQLESSDILLDFEV--LVFGKDG-----EKRRHEKLTSSIEYKRGHGDYA 55
Db 33 KVRDNGFVPHLYAKNKDLYEAYGYVMAKDRLFQLEMPFRGNEGTVSEI-----FGEDYL 88
QY 56 -----KYAERTABELQYGSNSFASFIKGEVLYKHEILCDVCDKLVKNYN-- 100
Db 89 SKDEQSRDGYSNKEIKKMDGLDROPKELLIAFAEGISRYVNEALKDPDDKLSKEFHEY 148
QY 101 -----KKTETTLIEQNMLS-----KILERSLEEMDDBEVKEMCDELSTKN- 140
Db 149 QFLPQKWTSTDVVRVYVMSVNTYPMFNKHQELAKNAEILAKLEHEYGTEVSRKXCFDDLVWKND 208
QY 141 -----TDNLNRQALSAATLTLFKMGFK-----SYQLAVIVANAVAK 177
Db 209 PSAPTIVSEGGPKRDSQSLQILSSAVIKASEKVKGRNENFVQTSSEELGLPLKTSNA 268
QY 178 TILGRGLSLAGNOVLRTLSFLTGP-VGWIITGVWTAIDIAIPAY 221
Db 269 AIWGEKSEKATGNALL-----FSGPQGVFVAPGFLYEVLHAPGF 307
```

Search completed: July 5, 2004, 03:12:51  
Job time : 12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:06:48 ; Search time 29 Seconds  
(without alignments)  
2752.624 Million cell updates/sec

Title: US-09-732-091-4

Perfect score: 1279

Sequence: 1 MAYKYDRDLFLKQLSSDL.....LRLKTTQANGKKSILQIESI 253

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 100%

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL 25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_rhiz.\*  
8: sp\_organelle.\*  
9: sp\_plant.\*  
10: sp\_rodent.\*  
11: sp\_virus.\*  
12: sp\_vertebrate.\*  
13: sp\_unclassified.\*  
14: sp\_rv.\*  
15: sp\_rv.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	265	20.7	237	16 Q8XGV3	Q8XGV3 salmonella
2	260	20.3	237	16 Q81SR4	Q81SR4 shigella fl
3	259	20.3	237	16 Q8FLC7	Q8FLC7 escherichia
4	133	10.4	39	16 Q26108	Q26108 helicobacte
5	105	8.1	1245	5 Q81ANI	Q81ANI plasmodium
6	103.5	8.1	403	16 Q8R9Z4	Q8R9Z4 thermoanaer
7	102.5	8.0	276	16 Q8D1H4	Q8D1H4 synecococc
8	102	8.0	1590	5 Q8MX30	Q8MX30 naegleria g
9	101	7.9	1455	5 Q8IKG8	Q8IKG8 plasmodium
10	100.5	7.9	726	13 P87397	P87397 oncorhynch
11	99.5	7.8	556	16 Q7VHM0	Q7VHM0 helicobacte
12	99.5	7.8	633	16 Q882L0	Q882L0 pseudomonas
13	99	7.7	202	5 Q81JZ2	Q81JZ2 plasmodium
14	99	7.7	222	4 O43633	O43633 homo sapien
15	99	7.7	722	11 Q9DB34	Q9DB34 mus musculu
16	99	7.7	723	13 Q7ZZU9	Q7ZZU9 astyanax fa

17	98.5	7.7	1009	16 Q99UD0	Q99UD0 staphylococ
18	98	7.7	282	16 Q8ZCL9	Q8ZCL9 vercinia pe
19	98	7.7	413	9 Q8SDW1	Q8SDW1 bacterioph
20	98	7.7	1099	16 Q98QT2	Q98QT2 mycoplasma
21	98	7.7	2359	5 Q81297	Q81297 plasmodium
22	98	7.7	2849	5 Q81HY4	Q81HY4 plasmodium
23	97.5	7.6	311	16 Q9PIC9	Q9PIC9 campylobact
24	97.5	7.6	728	16 Q9PI94	Q9PI94 campylobact
25	97.5	7.6	792	10 Q9ARQ9	Q9ARQ9 oryza sativ
26	96.5	7.5	1102	16 Q827Q7	Q827Q7 streptomyce
27	96.5	7.5	1341	5 Q81EQ5	Q81EQ5 plasmodium
28	96	7.5	262	5 Q61768	Q61768 caenorhabdi
29	96	7.5	420	16 Q97MM5	Q97MM5 cistridium
30	96	7.5	4405	5 Q81LZ2	Q81LZ2 plasmodium
31	95.5	7.5	944	17 Q87PC2	Q87PC2 methanosarc
32	94.5	7.4	406	16 Q8RH57	Q8RH57 fusobacteri
33	94	7.3	385	2 Q92188	Q92188 borrelia bu
34	94	7.3	576	10 Q84LB2	Q84LB2 malus domes
35	93.5	7.3	1413	5 Q81589	Q81589 plasmodium
36	93	7.3	355	17 Q97UF7	Q97UF7 sulfobus
37	93	7.3	610	16 Q51577	Q51577 borrelia bu
38	93	7.3	1173	2 Q9KHU3	Q9KHU3 helicobacte
39	92.5	7.2	391	9 Q7YSE5	Q7YSE5 aeromonas p
40	92.5	7.2	430	5 Q61767	Q61767 caenorhabdi
41	92.5	7.2	1960	10 Q8S6M2	Q8S6M2 oryza sativ
42	92	7.2	308	16 Q9PPG7	Q9PPG7 campylobact
43	92	7.2	309	16 Q8E6J7	Q8E6J7 streptococ
44	92	7.2	309	16 Q8E143	Q8E143 streptococ
45	92	7.2	331	17 Q96XT6	Q96XT6 sulfobus

## ALIGNMENTS

RESULT 1	Q8XGV3	PRELIMINARY;	PRT;	237 AA.
ID	Q8XGV3			
AC	Q8XGV3			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DE	Hypothetical protein (Positive regulator for sigma H (Sigma 32)			
DE	promoters, permitting growth at high temperature).			
GN	T0010 OR H1GA OR STM0010 OR STY0010.			
OS	Salmonella typhi, and			
OS	Salmonella typhimurium.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=601, 602;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			
RT	"Comparative Genomics of Salmonella enterica serovar Typhi strains Ty2			
RT	and CT18.";			
RL	J. Bacteriol. 185:2330-2337(2003).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RT	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856 (2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			

```

RX MEDLINE=21534947; PubMed=11677628;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Cornerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogg A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G., of a multiple drug resistant Salmonella
RT enterica serovar Typhi Ctr18.*;
RL Nature 413:848-852(2001).
DR EMBL; AEO16834; AAO67744.1; -
DR EMBL; AEO08693; AAL18974.1; -
DR EMBL; AL627265; CAD01163.1; -
DR InterPro; IPR005367; UPF0174.
DR Pfam; PF03667; UPF0174; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 237 AA; 26515 MW; 184ADE026EA5BCA9 CRC64;

Query Match 20.7%; Score 265; DB 16; Length 237;
Best Local Similarity 32.8%; Pred. No. 2.6e-12;
Matches 79; Conservative 40; Mismatches 101; Indels 22; Gaps 7;

QY 1 MAYKYDRLEFLKQLESSDLDLFEVLVFGKDGK-----HNEKLTSSIEYKRGDDY 54
DB 3 VYTHDEDLDFLQHCSEELQADFAILLTHNEKGKARLLSVLSHNE-LFKAME--GHPEQH 59
QY 55 AKYAEIABELOYGNSPASIKGEGVLYKEILCDVCKLVNKKYKTTTLIEQNMLS 114
DB 60 RENWQIIAGEFOHYGGDSIANLRGKGQYRAILLDVAKRLKXADKSMSTFEIEQQLLE 119
QY 115 KILERSLEEMDDEEVKEMCDELSIKNTNINQALSAATLTFK-MGGFKSVOLAVIVAN 173
DB 120 HFLRWTKQDAHQEFLQAVDAKVSELELLPLMKDRLSKAGVSHLSLTQLRIURT 179
QY 174 AVAKTILGRGLSLAGNOVLTRTSLFTGPGVMIITGVWTAIDIAGPAYRVTPACTVWAT 233
DB 180 HAAMSLILGHGL-LRG-----AGLGPGVGAALNGVKA--MSSGAYRVTPAVLIQIAC 227
QY 234 LR 235
DB 228 LR 229

RESULT 2
Q83SF4
ID Q83SR4 PRELIMINARY; PRT; 237 AA.
AC Q83SR4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative oxidoreductase.
GN SF0012 OR S0012.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jir Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
RW [2]
SQ SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;

```

```

RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AEO15039; AAN41678.1; -
DR EMBL; AEO16978; AAP15557.1; -
DR InterPro; IPR005367; UPF0174.
DR Pfam; PF03667; UPF0174; 1.
KW Complete proteome.
SQ SEQUENCE 237 AA; 26709 MW; 95509A1BCB8B4CF5 CRC64;

Query Match 20.3%; Score 260; DB 16; Length 237;
Best Local Similarity 30.8%; Pred. No. 6.3e-12;
Matches 76; Conservative 44; Mismatches 91; Indels 36; Gaps 6;

QY 3 YKYDRLEFLKQLESSDLDLFEVLVFGKDGKRNHEKLTSSIEYKR--HGDDYAKYAE 59
DB 5 YLNDSDLDLQHCSEELQANFARLLTHNEKGKTRLSILNRNELPKSMGHPQCHRRNQ 64
QY 60 RIAELOYGNSPASIKGEGVLYKEILCDVCKLVNKKYKTTTLIEQNMLSKILER 119
DB 65 LIAGLQHFQGGDSIANLRGKGKLYRAILLDVSKRLKADKEMSTFEIEQQLLEQLRN 124
QY 120 SLEEMDDE-----EVKEMCDELSIKNTNINQALSAATLTFKMGFKSVOLA 168
DB 125 TWKQDEEHKQEFHVAVDARVNELELLPLMKDKLLAGVS-----HLLSSQLT 174
QY 169 VIVANAVAKTILGRGLSLAGNOVLTRTSLFTGPGVMIITGVWTAIDIAGPAYRVTPAC 228
DB 175 RILRTHAAMSVILGHGL-LRG-----AGLGPGVGAALNGVKA--VSGSTRVTPAV 222
QY 229 IVVATLR 235
DB 223 LQIACLR 229

RESULT 3
Q8FLC7
ID Q8FLC7 PRELIMINARY; PRT; 237 AA.
AC Q8FLC7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created);
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein yaaw.
GN YAAW OR C0016.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22398234; PubMed=12471157;
RA Weich R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasiko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AEO16755; AAN78516.1; -
DR InterPro; IPR005367; UPF0174.
DR Pfam; PF03667; UPF0174; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 237 AA; 26727 MW; B8C190712375B31D CRC64;

Query Match 20.3%; Score 259; DB 16; Length 237;
Best Local Similarity 30.8%; Pred. No. 7.5e-12;
Matches 76; Conservative 44; Mismatches 91; Indels 36; Gaps 6;

```





Db 113 ASDFEVREKLEVALQ-----GPIKERNFTKEIAKKSIDELKENFNELSPVIER 164  
QY 111 NM-LSKILERSLEEMDEDEKEMCELSIKNTD-----NUNROALSAA----- 152  
Db 165 HIKLAKVKLODKPFLQDLNOKQFAEENVQKQDKTSQKIDKEAFTAAKEBEKTEKS 224  
QY 153 -----TLTIFKMGKFSYQLAVTAVANAVAKTILGRGLSLAGNQLVRLTSLFLACPGVMI 207  
Db 225 EDVKQEFVFFKNEG---KPSNLTYSIKS-----NDFVRLFR-----QIV 264  
QY 208 TGVWTAIDAGPAYRTIPACI-----VVATL-----RLKTOQ 240  
Db 265 DNVFVAKERKASSVTNMLKPEILGLKQLISLKSIDGNIVATVTESEKTKQIESNLSLQ 324  
QY 241 ANGDKKSQTESI 253  
Db 325 AQDLKGIKIESV 337

## RESULT 7

Q8DIH4 PRELIMINARY; PRT; 276 AA.  
AC Q8DIH4;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE T111614 protein.  
GN T111614.  
OS Synchococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.  
CX NCBI\_TaxID=32046;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BP-1;  
EX MEDLINE=22225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shingo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
Thermosynechococcus elongatus BP-1.";  
RL DNA Res. 9:123-130(2002).  
DR EMBL; AF065374; BAC09166.1; --.  
KW Complete proteome.  
SQ SEQUENCE 276 AA; 31282 MW; D0577A3D97E0CA92 CRC64;

Query Match 8.0%; Score 102.5; DB 16; Length 276;  
Best Local Similarity 19.0%; Pred. No. 7.2;  
Matches 50; Conservative 53; Mismatches 111; Indels 49; Gaps 10;  
QY 12 LKQLESSDLDLFEVLVFGKDGKRN--EKLTSSTIEYKRGDDYAKVAERIAEELQYVG 69  
Db 24 LELATEEELQDLTEILF-----RRRLNPDIYLTTPDPAVQDQRAWLDDIEERFLA 78  
QY 70 SNSFASFTKGE--VLYKEILLCDVCKLKVNNKTTTTLIBONMLSKILERSLEEMDE 127  
Db 79 ADGL-TVLKGRQAQISYRQTLMRVCYRLKIKFSPSWPELEIFNLVLRQMKKIGDQ 137  
QY 128 EVKEMCELSIKNTDNLNRQALSATLTFLKMGKFSYQLAVIVANAVAKTIL----- 180  
Db 138 DRRVLAQIQESLPDLHGHPISEMENRVLVEGG-----AAAISSVVRVMVQVQVARQ 191  
QY 191 -----GRGLSA-----GNQVLTSLTFLTGPVGMITGVWTAIDAGPA--- 220  
Db 192 FAIRPAGSKLSIAPLVSRAAGVARLAVGRSLAFVSTALW-----VWFIADLGHQALST 247  
QY 221 -YRTIPACIVATLR-LKTOQA 241  
Db 248 NYARIIEP-IFAIAQIRLLRGEQA 270

## RESULT 8

Q8MX30

ID Q8MX30 PRELIMINARY; PRT; 1590 AA.  
AC Q8MX30;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE RNA polymerase II largest subunit (Fragment).  
GN RPB1.  
OS Naegleria gruberi.  
OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.  
CX NCBI\_TaxID=5762;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=12032239;  
RA Dacks J.B., Marinets A., Ford Doolittle W., Cavalier-Smith T.,  
RA Logsdon J.M. Jr.;  
RT "Analyses of RNA Polymerase II Genes from Free-Living Protists:  
Phylogeny, Long Branch Attraction, and the Eukaryotic Big Bang.";  
RL Mol. Biol. Evol. 19:830-840(2002).  
DR EMBL; AF395110; AAM45151.1; --.  
DR GO; GO:0005665; C:DNA-directed RNA polymerase II, core complex; IEA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0004366; P:transcription from Pol II promoter; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR InterPro; IPR006592; RNA polA N.  
DR InterPro; IPR006592; RNA polA N.  
DR InterPro; IPR00684; RNA polII repeat.  
DR InterPro; IPR00722; RNA pol A.  
DR InterPro; IPR007080; RNA pol Rpb1\_1.  
DR InterPro; IPR007066; RNA pol Rpb1\_3.  
DR InterPro; IPR007083; RNA pol Rpb1\_4.  
DR InterPro; IPR007081; RNA pol Rpb1\_5.  
DR InterPro; IPR007075; RNA pol Rpb1\_6.  
DR InterPro; IPR007073; RNA pol Rpb1\_7.  
DR Pfam; PF04997; RNA pol Rpb1\_1; 1.  
DR Pfam; PF00623; RNA pol Rpb1\_2; 1.  
DR Pfam; PF04983; RNA pol Rpb1\_3; 1.  
DR Pfam; PF05000; RNA pol Rpb1\_4; 1.  
DR Pfam; PF04998; RNA pol Rpb1\_5; 1.  
DR Pfam; PF04992; RNA pol Rpb1\_6; 1.  
DR Pfam; PF04990; RNA pol Rpb1\_7; 1.  
DR SMART; SM00501; RNA pol Rpb1\_R; 5.  
DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; 7.  
FT NON\_TER 1590 1590  
FT SEQUENCE 1590 AA; 177446 MW; 5F4C7959D4D77D68 CRC64;  
Query Match 8.0%; Score 102; DB 5; Length 1590;  
Best Local Similarity 22.8%; Pred. No. 73;  
Matches 61; Conservative 37; Mismatches 71; Indels 98; Gaps 14;  
QY 28 VFGKDGKRNKLTSSIEYKRGDDYAKVAERIAEELQYVG--NSFASF-----I 77  
Db 810 LYGEDGMDAHKVE--SQSIDIMNLS--AKFDK-----YWNLSGSSPSFTPTSSVSM 861  
QY 78 KGEVJ-----YKHLCDVCKL-KVNNKKT-----ETTLIBON 111  
Db 862 PSGGDKESFSITKVYDEILKDPATYLSKVRBEYKTLMEDRHILRTEIPPNAENKIVMV 921  
QY 112 MLSKILERSLEEMD-----DEEVKEMCELSIKNTDNLNRQALSATLT 154  
Db 922 NLKRIIKNAQKEYGHPSIGKPSDMNPYIVIEKIKLCCDLIVVTKGDELSTEAQTNATL 981  
QY 155 TLFKMGKFSYQLAVIVANAVAKTILGRGLSLAGNQLVRLTSLTGPVGMIT----- 207  
Db 982 -LFSM-----ELRSTFFPKGCSLKNCVLLKXPLTF-----NLVKLELRPH 1020  
QY 208 -----TGVWTAIDAGPAYRTI 225  
Db 1021 NCLAQPGEMVGSVAQAQSIGEPATQMTL 1047

```

RESULT 9
Q8IKG8
ID Q8IKG8 PRELIMINARY; PRT; 1455 AA.
AC Q8IKG8
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DI 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Rhoptry protein, putative.
GN PF14_0637.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Carlton M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
Eisen J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
Chan M.-S., Rene V., Shallow S.J., Suh B., Peterson J., Anguoli S.,
Perteira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
Martin D.M.A., Fairclamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
McFadden G.I., Cummings L.H., Subramanian G.M., Mungall C.,
Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
Fraser C.M., Barrell B.
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014826; AAN37250.1; --
GO GO:0005351; F:sugar porter activity; IEA.
GO GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.
DR InterPro; IPR002114; HPR Serp S.
DR PROSITE; PS00589; PTS_HPR_SER1.
SQ SEQUENCE 1455 AA; 174305 MW; 36FPA1760A898F99 CRC64;

Query Match 7.9%; Score 101; DB 5; Length 1455;
Best Local Similarity 25.5%; Pred. No. 78;
Matches 36; Conservative 29; Mismatches 52; Indels 24; Gaps 6;

QY 6 DRDLFLKQLSSDLLDLFEVLVFGKDEKHEKLTSSIEYKHGDDYAKYAEIAE 64
Db 333 ERELYIYGLME-----ELRNEIKTKEEGNNIKLENKIEYKQNEELRNEKEKLQST 387
QY 65 LQYGSNSFASFIKGEVGLVYKILCDVCKLKNVY---KXTETLLQNNMLSKILRESL 121
Db 388 INEY-SHNFNLDHDKITNKE-----CEELKNNYTIKELYELKKEQEIYK----- 435
QY 122 EEMDDEEVKEMCELSTANTD 142
Db 436 ---QEEYKSLDELENNNE 453

RESULT 10
P87397
ID P87397 PRELIMINARY; PRT; 726 AA.
AC P87397
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DI 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Heat shock protein hsp90.
OS Oncorhynchus tshawytscha (Chinook salmon) (King salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=74940;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99262176; PubMed=10329464;
RA Palmisano A.N., Winton J.R., Dickhoff W.W.;
RT "Sequence features and phylogenetic analysis of the stress protein
hsp90alpha in chinook salmon (Oncorhynchus tshawytscha), a
poikilothermic vertebrate."
RL Biochem. Biophys. Res. Commun. 258:784-791(1999).

```

```

DR EMBL; U89945; AAB49983.1; --
DR HSSP; P07900; lBYQ.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR001404; Hsp90.
DR Pfam; PF02518; HATPase_C; 1.
DR Pfam; PF00183; HSP90; 1.
DR PRINTS; PRO0775; HEATSHOCK90.
DR SMART; SM00387; HATPase_C; 1.
DR PROSITE; PS00298; HSP90; 1.
SQ SEQUENCE 726 AA; 83505 MW; 1A2A962C471C8CF2 CRC64;

Query Match 7.9%; Score 100.5; DB 13; Length 726;
Best Local Similarity 21.6%; Pred. No. 35;
Matches 61; Conservative 39; Mismatches 81; Indels 101; Gaps 11;

QY 21 LDLEEVVFGKDEKHEKLTSSIEYKHGDDYAKYAEIAEIQYGSNSFASFIKGE 80
Db 414 MDLFVELSEDKNYKFKYEQFSKNIKLGIHED--AQNRKKLSMDLRYTTSNAD---E 467
QY 81 GVLVYKILCDVCKLKNVYKXTET-----LSE----- 109
Db 468 MVLSEKYSVRMKDTQKHLIYITGETKEQVANSFVERLRKAGLEVIYMIPIDEYCVQQL 527
QY 110 -----QNNMLSKILERSLEEMDDEEVKEMCELSIKNTNLNQAALTLFLKMGGRKS 164
Db 528 KEYDGNLVS-VTKEGLELPEDDEKXKQELNFK-FENLCK----- 567
QY 165 YQLAVIVANAKTILGRGLS--LAGNQVLRITLSFLGPGVWILIT----- 208
Db 568 -----TKDILDKKIEKVSYSNLVSPCCIVTSTYGTWTANMERIMKSQLRDN 617
QY 209 --GVWTA---IDIAGPAYRVTPACIVVATLRKTKQOANGDK 245
Db 618 TWGYNTAKKHLIN-----PTHPIVETLREKABADKNDK 651

RESULT 11
Q7VHM0
ID Q7VHM0 PRELIMINARY; PRT; 556 AA.
AC Q7VHM0
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DI 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN HH0946.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
Helicobacter hepaticus."
OC Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR EMBL; AE017147; AAF77543.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 556 AA; 61984 MW; 0E48CF5376BAC48D CRC64;

Query Match 7.8%; Score 99.5; DB 16; Length 556;
Best Local Similarity 23.8%; Pred. No. 30;
Matches 53; Conservative 33; Mismatches 104; Indels 33; Gaps 6;

QY 32 DGEKRHEKLTSSIEYKHGDDYAKYAEIAEIQYGSNSFASFIKGEVGLVYKILCDV 91
Db 199 DQYKRHEKLTSSIEYKHGDDYAKYAEIAEIQYGSNSFASFIKGEVGLVYKILCDV 243

```

```
QY 92 CDKLKYNVKKTTTL-----IQNMLSKILERSLEEMDDE--EVKEMCD-----E 135
Db 244 CNRLMCENPRITAVATQSIWASGHIAQAGFSDAIIAALATLANGVIWVEXKMFEGSIDTE 303
QY 136 LSI-KNTDNLNRQALSAATITLFKMGFGKSYQALAVIVANAVAKTILGRGLSLAGNOVLTR 194
Db 304 TSILKRIKLLKRTIEAFQTFGRGAGGALDAVGVGVQIFRSIAGK-LKLVWDKIRTA 362
QY 195 TSLFLGPGVCIITGWTAIDIAGPAVRVTIPACIVVATURLK 237
Db 363 LKSIYNGIVGIKGEVSNRELIIILKSLFSAWVWSTLALE 405

RESULT 12
Q882L0 PRELIMINARY; PRT; 633 AA.
AC Q882L0;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DE MethyL-accepting chemotaxis protein.
GN PSPT02616.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Uterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016865; AAC056120.1; -.
DR TIGR; PSPT02616; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0034871; F:signal transducer activity; IEA.
DR GO; GO:0003935; F:chemotaxis; IEA.
DR GO; GO:00037165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chntaxis_trans.
DR InterPro; IPR033660; HAMF.
DR Pfam; PF00672; HAMF; 1.
DR Pfam; PF00015; MCPsigal; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMF; 1.
KW Complete proteome.
SQ SEQUENCE 633 AA; 68090 MW; D15A8BC00AC58637 CRC64;

Query Match 7.8%; Score 99.5; DB 16; Length 633;
Best Local Similarity 23.7%; Pred. No. 35;
Matches 69; Conservative 50; Mismatches 87; Indels 85; Gaps 18;

QY 3 YKXDRLEFLKQLESSDLIDLFEVLVFGKDGKRNKLTSS-----IEYKRGDD---- 53
Db 70 YDAERASDWLKALEAHVHVAQEVF---DSPNIPLVNAAEALVEYRVHYDNLMPA 125
QY 54 -----YAKYAEIAEELQYVGSNPFASFKGEG-----VLYKEILCDWCD 93
Db 126 TAAREATRGAFQYADAGAELOKL--NAFARSDEGSASORDAIVQAMTLFQKMFCL-- 181
QY 94 KLVKNVKKTTTLTIQNMLSKILB--RSLEEMDDEE--VKEMCDL-SIKNTDN---LN 145
Db 182 -RGYTSYKAEENRAPAESMSAVINVFVGLQGFQDSATIKHLVDMSVSYQNTMNFQATA 240
QY 146 RQALSAATLITFKMGG--FKSY-QLAVIVANAV-----AKTILGRGLSLAGNOVLTR 193
Db 241 QASIDQAQAGITKVGVLFKSADQLS---ANQVSLRIEDVDQARTLLSVWL-----VAA 291
QY 194 RTLFLGPGVCIITGWTAIDIAGPAVRVTIPACIVVATURLK 244
```

```
Db 292 LIMSFL---AAWVITRL-----IVGP-----LLETILKLAERVAQGD 324

RESULT 13
Q81JZ2 PRELIMINARY; PRT; 202 AA.
AC Q81JZ2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN Pf10_0048.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RC MEDLINE=2255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chant M., Kene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Martin D.M.A., Fairlamb A.H., Fraumholz M.J., Roos D.S., Vaidya A.B.,
RA McFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Nature 419:498-511(2002).
DR EMBL; AB014829; AAN35246.1; -.
KW Hypothetical protein.
SQ SEQUENCE 202 AA; 24259 MW; 48FD22A7F21165A6 CRC64;

Query Match 7.7%; Score 99; DB 5; Length 202;
Best Local Similarity 24.9%; Pred. No. 8;
Matches 46; Conservative 31; Mismatches 50; Indels 58; Gaps 9;

QY 4 KYDRDLEFLKQLESSDLIDLFEVLVFGKDGKRNKLTSS-----EKL--TSSIEYKRGDDYAK 56
Db 32 KYEENILLLKLMRTENIKF-----DQYNAINILEKLDIEDIQYEDILNDYAK 81
QY 57 -----YARIABEELQYVGSNPFASFKGEGVLYKILCDVCDKLKYNVKK----- 102
Db 82 NFLTLKSGSFKVAEEIQI---NDKISFRNE--LIKQDLNNKLQNLNHNLYNLYNLNK 136
QY 103 ---TE-----TTLIEQNMLSKILERSLEEMDDEEVK-----EMCDELSIKN 140
Db 137 CLGTDFESQINASIKFNSKILSKLKKEMKDYEQYNNLSDTDFELDMKICELENYLN 196
QY 141 TDNLN 145
Db 197 NTNQN 201

RESULT 14
Q43633 PRELIMINARY; PRT; 222 AA.
AC Q43633;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DE BC-2 protein (putative breast adenocarcinoma marker) (32KD) (32
DE kDa).
GN BC-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: July 5, 2004, 03:07:28 ; Search time 16 Seconds  
(without alignments)  
816.336 Million cell updates/sec

Title: US-09-732-091-4  
Perfect score: 1279  
Sequence: 1 MAYKYDRLEFLKQLESSDL.....LRLKTOQANGKSLQIHESI 253

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgm2\_6/ptodata/2/iaa/5A COMB.pcp.\*  
2: /cgm2\_6/ptodata/2/iaa/5B COMB.pcp.\*  
3: /cgm2\_6/ptodata/2/iaa/6A COMB.pcp.\*  
4: /cgm2\_6/ptodata/2/iaa/6B COMB.pcp.\*  
5: /cgm2\_6/ptodata/2/iaa/PCTUS COMB.pcp.\*  
6: /cgm2\_6/ptodata/2/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	364	28.5	273	4	US-09-252-991A-29841 Sequence 29841, A
2	346.5	27.1	321	4	US-09-252-991A-29840 Sequence 29840, A
3	247	19.3	258	4	US-09-489-039A-8852 Sequence 8852, A
4	100	7.8	739	4	US-09-543-681A-6437 Sequence 6437, A
5	99	7.7	222	3	US-08-944-604-16 Sequence 16, A
6	99	7.7	274	4	US-09-134-001C-5279 Sequence 5279, A
7	96	7.5	173	3	US-08-658-639-12 Sequence 12, A
8	96	7.5	173	3	US-08-944-604-12 Sequence 12, A
9	95	7.4	2125	4	US-09-919-172-29 Sequence 29, A
10	91.5	7.2	1086	4	US-09-543-681A-7696 Sequence 7696, A
11	89.5	7.0	281	4	US-09-198-452A-749 Sequence 749, A
12	89.5	7.0	543	4	US-09-328-352-5845 Sequence 5845, A
13	89.5	7.0	546	4	US-09-345-236B-98 Sequence 98, A
14	89.5	7.0	546	4	US-09-345-236B-121 Sequence 121, A
15	89.5	7.0	560	4	US-09-446-301A-50 Sequence 50, A
16	88.5	6.9	552	4	US-09-446-301A-4 Sequence 4, A
17	88.5	6.9	552	4	US-09-099-932-A Sequence 4, A
18	88	6.9	488	4	US-09-489-039A-13363 Sequence 13363, A
19	87.5	6.8	732	2	US-08-533-669A-18 Sequence 18, A
20	87.5	6.8	732	4	US-09-307-143-4 Sequence 4, A
21	87.5	6.8	732	4	US-09-183-861-18 Sequence 18, A
22	87.5	6.8	732	4	US-09-022-765-18 Sequence 18, A
23	87.5	6.8	732	4	US-09-551-974A-18 Sequence 18, A
24	87.5	6.8	732	4	US-09-565-501A-18 Sequence 18, A
25	87.5	6.8	732	4	US-09-619-206A-18 Sequence 18, A
26	87.5	6.8	732	4	US-09-874-923-18 Sequence 18, A
27	87	6.8	491	3	US-09-029-267-2 Sequence 2, A

28	86.5	6.8	331	4	US-09-134-001C-3626 Sequence 3626, A
29	84.5	6.6	528	4	US-08-637-670-27 Sequence 27, A
30	84.5	6.6	691	4	US-09-830-217-12 Sequence 12, A
31	84	6.6	492	4	US-09-107-532A-7062 Sequence 7062, A
32	83.5	6.5	264	2	US-08-719-697-8 Sequence 8, A
33	83.5	6.5	264	3	US-08-727-616A-8 Sequence 8, A
34	83.5	6.5	264	4	US-09-481-756-8 Sequence 8, A
35	83.5	6.5	286	1	US-07-721-775A-2 Sequence 2, A
36	83.5	6.5	286	1	US-08-339-658-2 Sequence 2, A
37	83.5	6.5	286	3	US-09-263-933-7 Sequence 7, A
38	83.5	6.5	286	3	US-09-263-933-14 Sequence 14, A
39	83.5	6.5	286	3	US-09-025-769B-265 Sequence 265, A
40	83.5	6.5	286	4	US-09-025-769B-362 Sequence 362, A
41	83.5	6.5	286	4	US-09-919-901-7 Sequence 7, A
42	83.5	6.5	286	4	US-09-919-901-14 Sequence 14, A
43	83.5	6.5	286	4	US-09-919-901-21 Sequence 21, A
44	83.5	6.5	286	4	US-09-025-769B-285 Sequence 285, A
45	83.5	6.5	299	4	US-09-025-769B-285 Sequence 285, A

ALIGNMENTS

RESULT 1  
US-09-252-991A-29841  
; Sequence 29841, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29841  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29841

Query Match 28.5%; Score 364; DB 4; Length 273;  
Best Local Similarity 35.4%; Pred. No. 1.1e-30;  
Matches 85; Conservative 49; Mismatches 88; Indels 18; Gaps 4;

QY	10	EFLKQLESSDLDLDFVLVFGKDGKRRNEKLTSTSEYKXHGDDYAKYAEIRIABELQYVG 69
DB	28	QLLERSVNDLFLVEYIL-----KARTESLSKQVDFKQWHPHRRVYSAIDLRLFG 81
QY	70	SNSFASFIKGGVLYKEILCDVCDKLVNKKTKTTLTLEONMLSKILERSLEMDDEEV 129
DB	82	GNSEALWRKSGPSYTEVVRDVAGLKVGVGSMELIESEAMVQSILQKXSGEDR 141
QY	130	KEMCDEL-----SINKTDLNARQALSAATLTFKMGPKSYCLAVIVANAVAKTILGRG 183
DB	142	RELEETIRLAGLDTKTKTALLNGSLSG--LVFVAVARMILYKTSIVVNSMAQLLCHG 199
QY	184	LSLA----GNQVLTSLTFLTPGVGWIITGVWTAIDAGPAYRVVTPACIVVATRLKQT 239
DB	200	LRSNVVAGTTFAGGRAVALAGPVGVVAGVWTVADLAGPAYRVVTPICVLHIAMLRLKAR 259

RESULT 2  
US-09-252-991A-29840  
; Sequence 29840, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252.991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 29840  
;; LENGTH: 321  
;; TYPE: PRT  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29840

Query Match 27.1%; Score 346.5; DB 4; Length 321;  
Best Local Similarity 34.3%; Pred. No. 1.1e-28;  
Matches 82; Conservative 42; Mismatches 112; Indels 3; Gaps 3;

Qy 1 MAYKYRDL-EFLKQLESSDLDFEVLVFGKGEKRNKLTSSIEYKRGHDDYAKYAE 59  
Db 53 MAITHDADLADVLSSDDIRLLDVIDTNGRISLSSVCQKLSAAKGVVGFEPERG 112  
Qy 60 RIABELQYGSNSPASFPIK-GEGLVYKEILCDVCKLKVYNNKKTETTLIEQNMLSKILE 118  
Db 113 MYAEELMRFGNSLNNLFRGGSGVYFYKELLSDVASHVGSVKESTGDCARMEMALITKVFE 172  
Qy 119 RSLSEMDDEEVKEMCDELSIKNTDNLNRQALSAATLILFKMGGFKSQVLAVIVANAVAKT 178  
Db 173 QSIGRSEEDKATFEESIGTSYRSGMGFVAL-ALIASLSASGWTSGYLAAMVASATWSS 231  
Qy 179 ILRGSLAGNQVLTSLFUTGPVGMIIITGWTATDIAGPAYRVTTIPACIVATLRLK 237  
Db 232 LVGRGVALAGCATLGRGLSVLTGPVGMIAIGTAFDIASPAYRVTLPCVQIGHMRQK 290

RESULT 3  
US-09-489-039A-8852  
;; Sequence 8852, Application US/09489039A  
;; Patent No. 6610836  
;; GENERAL INFORMATION:  
;; APPLICANT: Gary Breton et. al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
;; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
;; CURRENT APPLICATION NUMBER: US/09/489.039A  
;; CURRENT FILING DATE: 2000-01-27  
;; PRIOR APPLICATION NUMBER: US 60/117,747  
;; PRIOR FILING DATE: 1999-01-29  
;; NUMBER OF SEQ ID NOS: 14342  
;; SEQ ID NO 8852  
;; LENGTH: 258  
;; TYPE: PRT  
;; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8852

Query Match 19.3%; Score 247; DB 4; Length 258;  
Best Local Similarity 28.2%; Pred. No. 3.4e-18;  
Matches 70; Conservative 48; Mismatches 100; Indels 30; Gaps 6;

Qy 3 XYDRELEFLKQLESSDLDFEVLVFGKGEKRNKLTSS---IEYKRGHDDYAKYAE 59  
Db 14 YLEDCALFLPECSEAHLEATRIILTHGNGKPRSLSTLLRNETFLAMEGHPERYRRNQ 73  
Qy 60 RIABELQYGSNSPASFPIKGEGLVYKEILCDVCKLKVYNNKKTETTLIEQNMLSKILER 119  
Db 74 LIAGELCHFGGDSIANTLRHKKFYRAILLVDCKFLKAKVKQLSTPQIEQQLLAHFLQH 133  
Qy 120 SLEMDDEEVKEMCDELSIKN--TDNL-----NRQALSAATLILFKMGGFKSQVLAVIV 171  
Db 134 SWNKLNAQKQFLAAVECRSHELDS-NAHLIRHKLSEGVTLTL-----DERLTAL 186  
Qy 172 ANAVAKTILGRGLAGNQVLTSLFUTGPVGMIIITGWTATDIAGPAYRVTTIPACIV 231

Db 187 RTHAAVSIVGHGLVRGAG-----LGGPLGAALNSVKA---VSGSAYRVTIPAVLHI 234  
Qy 232 ATLRLKQTQ 239  
Db 235 ACLRQMLQ 242

RESULT 4  
US-09-543-681A-6437  
;; Sequence 6437, Application US/09543681A  
;; Patent No. 6605709  
;; GENERAL INFORMATION:  
;; APPLICANT: GARY BRETON  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA  
;; FILE REFERENCE: 2709.1002-001  
;; CURRENT APPLICATION NUMBER: US/09/543.681A  
;; CURRENT FILING DATE: 2000-04-05  
;; PRIOR APPLICATION NUMBER: US 60/128,706  
;; PRIOR FILING DATE: 1999-04-09  
;; NUMBER OF SEQ ID NOS: 8344  
;; SEQ ID NO 6437  
;; LENGTH: 739  
;; TYPE: PRT  
;; ORGANISM: Proteus mirabilis  
US-09-543-681A-6437

Query Match 7.8%; Score 100; DB 4; Length 739;  
Best Local Similarity 25.6%; Pred. No. 0.089;  
Matches 54; Conservative 42; Mismatches 65; Indels 50; Gaps 10;

Qy 6 DRDLEFLKQLESSDLDFEVLVFGKGEKRNKLTSSIEYKRGHDDYAKYAEIAEEL 65  
Db 517 DKVSDFVSLVNSSLDGLKEL---GLD--KKIIBEMKNSIQ-----DKLKMATKILVTV 565  
Qy 66 QYGSNSPASFPIKGEGLVYKEILCDVCK-----LVYNNKKTETTL--IEQNMLS 114  
Db 566 VLFVAATLSPVITGPAM---KGISDANVKINSQIROLKVLNDGLAVLGMKMDIII 622  
Qy 115 KILERSLEEMDDEEVKEMCDELSIKNTDNLNRQAL-----SAATLILFKMGGFKSQYOLA 168  
Db 623 KALEALEKID---KQLAKEISKASIMLARVTVVASKLTNSAINTVNIYGS----- 671  
Qy 169 VIVANAVAKTILGRGLAGNQVLTSLF 199  
Db 672 -VIASKIQ-----SIAGSKLTAVIDII 694

RESULT 5  
US-08-944-604-16  
;; Sequence 16, Application US/08944604  
;; Patent No. 6218131  
;; GENERAL INFORMATION:  
;; APPLICANT: KEESEE, SUSAN  
;; APPLICANT: OSAR, ROBERT  
;; APPLICANT: WU, YING-YUE  
;; TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF  
;; NUMBER OF SEQUENCES: 24  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Testa, Hurwitz & Thibault  
;; STREET: 125 High St.  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02110  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/944,604
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MEYERS, THOMAS C
; REGISTRATION NUMBER: 36,989
; REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-944-604-16

```

```

Query Match 7.7%; Score 99; DB 3; Length 222;
Best Local Similarity 21.4%; Pred. No. 0.018;
Matches 39; Conservative 48; Mismatches 83; Indels 12; Gaps 5;

```

```

QY 6 DEDLEFLKQESSDLDFEVLVFGK-DGEKRNEXLTSIBYKRHGGDYAKYAEIAE 64
DB 30 DREORQKLETOEKIADIKKMAKQGMQMDAVRIMAKDLVTRRYVRKFLMRANIQAVSLK 89
QY 65 LQYGS-NSFASFIKG-----EGVLYKEILCDVCKLKVYNNKTTTTLIEQNMLSKILER 119
DB 90 IQTLKSNNSMAQAMKGVTKAMGTMRNLKLPQTKIMMEFERQAEIMDKMKNMNDIAID 149
QY 120 SLEEMDDDE-----VKEMDELSIKNTDNLNRQALSAATLTLFKMGGKSYQLAVIVANA 174
DB 150 ANGDEDEESDAVNSQVLDLGLSLTDELNLPTGGSLSV-AAGGKKAASALADA 208
QY 175 VA 176
DB 209 DA 210

```

```

RESULT 6
US-09-134-001C-5279
; Sequence 5279, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5279
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5279

```

```

Query Match 7.7%; Score 99; DB 4; Length 274;
Best Local Similarity 23.8%; Pred. No. 0.025;
Matches 62; Conservative 44; Mismatches 88; Indels 66; Gaps 14;

```

```

QY 3 YKDRDLFLKQLESSLDLDFEVLVFGKDGKGRHNEKLTSS-----IE 46
DB 4 YNNNDL-----FLILKFGDENKDTIETISREALIKOVVMTAARILLE 47
QY 47 YKRHGGDYAKYAEIAEELQYGSNSFASFIKGEGVLYKEILCDVCKLKVYNNKTT-ET 105
DB 48 SGAEGTRVEDTWTARIATKLGYPESNSFVTNTVIEFVLHNPAYPRL-----YRIKTRDT 100

```

## RESULT 7

```

US-08-658-639-12
; Sequence 12, Application US/08658639
; Patent No. 5914238
; GENERAL INFORMATION:
; APPLICANT: KEESEE, SUSAN
; APPLICANT: OBAR, ROBERT
; APPLICANT: WU, YING-JYE
; TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,639
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MEYERS, THOMAS C
; REGISTRATION NUMBER: 36,989
; REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-658-639-12

```

```

Query Match 7.5%; Score 96; DB 2; Length 173;
Best Local Similarity 23.3%; Pred. No. 0.027;
Matches 27; Conservative 33; Mismatches 46; Indels 10; Gaps 3;

```

```

QY 70 SNSPASFPIKG-----EGVLYKEILCDVCKLKVYNNKTTTTLIEQNMLSKILERSLEEMD 125
DB 47 NNSMAQAMKGVTKAMGTMRNLKLPQTKIMMEFERQAEIMDKMKNMNDIAIDAMGDEE 106
QY 126 DEE-----VKEMDELSIKNTDNLNRQALSAATLTLFKMGGKSYQLAVIVANA 176
DB 107 DEESDAVNSQVLDLGLSLTDELNLPTGGSLSV-AAGGKKAASALADA 161

```

## RESULT 8

```

US-08-944-604-12
; Sequence 12, Application US/08944604
; Patent No. 6218131
; GENERAL INFORMATION:

```

APPLICANT: KEESEE, SUSAN  
APPLICANT: OBAR, ROBERT  
APPLICANT: WU, YING-YVE  
TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF  
BREAST CANCER  
TITLE OF INVENTION: BREAST CANCER  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Testa, Hurwitz & Thibault  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,604  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MEYERS, THOMAS C  
REGISTRATION NUMBER: 36,989  
REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 173 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-944-604-12

Query Match 7.5%; Score 96; DB 3; Length 173;  
Best Local Similarity 23.3%; Pred. No. 0.027;  
Matches 27; Conservative 33; Mismatches 46; Indels 10; Gaps 3;  
Qy 70 SNFSPAFKIG---EGVLYKEILCDVCDKLVNKKTTTLLIQNMLSKILERSLEMD 125  
Db 47 NNSMAQMKGVTKAMGTNMQKLPQTKIMMEPERQAEIMDKBEMWMDAIDAMGDEE 106  
Qy 126 DEE-----VKEMCDLSIKNTDNLNQAALSAATLTFMGGFKSYQLAVIVANAVA 176  
Db 107 DEESDAVVSQVLDLGLSLTDELSNLPSTGSLSV-AAGGKAAEAASALADADA 161

RESULT 9  
US-09-919-172-29  
Sequence 29, Application US/09919172  
Patent No. 6673545  
GENERAL INFORMATION:  
APPLICANT: Paris, Mary  
APPLICANT: Turner, Christopher M.  
TITLE OF INVENTION: PROSTATE CANCER MARKERS  
FILE REFERENCE: PA-0036 US  
CURRENT APPLICATION NUMBER: US/09/919,172  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/222,469  
PRIOR FILING DATE: 2000-07-28  
NUMBER OF SEQ ID NOS: 102  
SOFTWARE: PERL Program  
SEQ ID NO 29  
LENGTH: 2125  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6673545 3774181CD1  
US-09-919-172-29

Query Match 7.4%; Score 95; DB 4; Length 2125;  
Best Local Similarity 25.4%; Pred. No. 1.5;  
Matches 46; Conservative 35; Mismatches 62; Indels 38; Gaps 9;  
Qy 2 AVKYDRDLFLK--QLESSDLLDF-----EVLVFGKDGKRRHNEKLTSSIEYKRRHGDYA 55  
Db 986 SFRDEKELERLQICQKSDHLKQFEXSHQQLQNIKAENNDKIQLNELEKSNECA 1045  
Qy 56 KYAERIAEEL--QYGSNSFASFKGEG---VLYKEILCDVCDKLVK----- 97  
Db 1046 EMLKQKVEELTRQNNETKLMQRIQAESENIVLEKQTIQORCEALKIQADGFKDQLRSTN 1105  
Qy 98 -NYNKTETTLIQNMLSKI--LERSLEEMDD--EYKEMCDELSI-----KNTDNLNR 146  
Db 1106 EHLEKQTKT---EQDFORKIKCLEEDLAKSONLVSEFKQKDDQONIIQNTKKEVRNLNA 1162  
Qy 147 Q 147  
Db 1163 E 1163

RESULT 10  
US-09-543-681A-7696  
Sequence 7696, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 7696  
LENGTH: 1086  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-7696

Query Match 7.2%; Score 91.5; DB 4; Length 1086;  
Best Local Similarity 19.5%; Pred. No. 1.3;  
Matches 43; Conservative 38; Mismatches 73; Indels 67; Gaps 9;  
Qy 3 YKYDRDLFLKQLESSDLLDFEVLVFGKDGKRRHNEKLTSSIEYKRRHGDYAKYAEIA 62  
Db 786 FAYDRGKE-IEOMQ-----FETTLIGK--SRAEQEKINA----- 816  
Qy 63 EELQYGSNSFASFKGEGVLYKEILCDVCDKLVNKKTTTLLIQNMLSKILERSLE 122  
Db 817 -----LRQIDVLYQOASVDLGEKELVNLQNVLT--KQIEELKREAM 860  
Qy 123 EMD-----DDEVKEMCDELSIKNTDNLNQAALSAATLTFMGGFKSYQLAVI 170  
Db 861 KGDPMAGLKGSLSDSESAMDVMEVNRVNTNALNMSDALADPALTKGSKFDPANAVI 920  
Qy 171 --VANAVAKTILGRGLSLAGNQLVTRTSLFUTGPGVCHITG 209  
Db 921 SDITRMVMKMLIFKAIEAGG-----QAMGF---DMGWMKSG 953

RESULT 11  
US-09-198-452A-749  
Sequence 749, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffais, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragme  
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, pr  
TITLE OF INVENTION: and treatment of infection  
FILE REFERENCE: 9710-003-999



```

; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 749
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-749

Query Match          7.0%; Score 89.5; DB 4; Length 281;
Best Local Similarity 22.0%; Pred. No. 0.27;
Matches 56; Conservative 49; Mismatches 103; Indels 47; Gaps 11;

Qy 22 DLFEVLVFGKDGKRNKELTSS-----IEYKRGDDYAKYARIAE-----ELQY 68
Db 30 LSRDVKVFLSEKQTRQETSSKGNTRTSERKFADEKRVDEIARVGSKEEEOESQEF 89
Qy 69 --GNSFEA--SPFKGEGVLYKEILCDVCKLKVYNNKKTETTLIEONMLSKILERSLEEM 124
Db 90 CLAENAFAGMLIDIAAGSRAEAVVEVA-PIAVS---SIDTQWIEIILSTVESMVISI 145
Qy 125 DDEEVKEMCDELSIKNTONLQALSAATLTILFKMGFKSYQLAVIVANAVAKTILGRGL 184
Db 146 NGEQVEL-----VLDASSVPEAFVGANLTLVQSG-----QLSVKFSFVDATQMAEAA 196
Qy 185 SLAGN--OVLTRTSLFLTG-----PVGWITGTWTAIDTAGPAYRVTIACIIVATL 234
Db 197 DLVTNPNQSLSLVSALSGHQLTLKEFSVGNLLVQLPKIEEYQTPLEH-----MIASIT 249
Qy 235 RLKTOQANGDKKSLQ 249
Db 250 RHREKQDQKQK 264

RESULT 12
US-09-328-352-5845
; Sequence 5845, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5845
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5845

Query Match          7.0%; Score 89.5; DB 4; Length 543;
Best Local Similarity 21.1%; Pred. No. 0.74;
Matches 40; Conservative 31; Mismatches 52; Indels 67; Gaps 10;

Qy 6 DEDLEFLKQLESLLDLEVLVFGKDGKRNKELTSSIEYKRGD-----DYA 55
Db 366 DKSLELEKEERKDLKNRFD-----EYNNKGQIKFVTFHOSFSYE 405
Qy 56 KYAERI-AEELQYGSNSPASFIFKGEGLVLYKEILCD-----VCDKLVYNNKKTETTLI 108
Db 406 DFEVIGRAETVEYSDGKNIEVPVSGVF--KLCDTAQSKVILESQKINFDSNT-----458
Qy 109 EQWMLSKILERSLEEM--DDEEVKEMC-----DELSIKNTDNLRQALSAATLT 156
Db 459 -----NEIWMKSLGRAGEDIDFYCIKNHCNVLLGMGDELDFNSA--VNRKQIE-----EI 507
Qy 157 FOMGFKFSYQ 166
Db 508 MDKNGYEAYR 577

; CURRENT APPLICATION NUMBER: US/09/345,236B
; Sequence 98, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tokuo, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 546
; TYPE: PRT
; ORGANISM: mosquito baculovirus
US-09-345-236B-98

Query Match          7.0%; Score 89.5; DB 4; Length 546;
Best Local Similarity 23.5%; Pred. No. 0.75;
Matches 60; Conservative 34; Mismatches 92; Indels 69; Gaps 10;

Qy 23 LFEVLVFGKDGKRNK-----LTSSIEYKRGDDYAKYAE---RIAELQYYSNS 72
Db 153 VFEVVGGEYGAAGEDEVSRSNLSLFDMASEVQSTDAAKVMELFSALSEQORNLNN 212
Qy 73 EASFIFKGEGLVLYKEILCDVCKLKVYNNKKTETTLIEQNMLSKILERSLEEMDDEVKEM 132
Db 213 FGAAFGSGGT-----TPPTSQDDM-----EVEDVETVEK 242
Qy 133 CDELSIKNTDNLN---RQALSAATLTILFKMG---GFKSYQLAVIVANAVAKTILGRGLS 185
Db 243 PENLNDIITDQLRDFMAQELKKAENYVFKMGSTVGESKSAIAITVADRVSRFSFMEGRI 302
Qy 186 LAGNOV-----LPRTISFTLPGVGNII-----TGWTAIDTAGPAYRVTI 227
Db 303 VDYNQVVLHILNDYDQLEBLLSFRKTK--YIIAEGVPHDSKVHYVDLT--QYRETVPY 358
Qy 228 CIVVATILRLKTOQAN 242
Db 359 SIALLNLSRGVDQAN 373

RESULT 14
US-09-345-236B-121
; Sequence 121, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tokuo, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 121
; LENGTH: 546
; TYPE: PRT
; ORGANISM: mosquito baculovirus
US-09-345-236B-121
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:06:13 ; Search time 39 Seconds  
(without alignments)  
1832.936 Million cell updates/sec

Title: US-09-732-091-4  
Perfect score: 1279  
Sequence: 1 MAYKYDRDLFLKQLLESSDL.....LRLKTKQANGKSKSLQIESI 253

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_25Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1279	100.0	253	5 AAE26860	Aae26860 Helicobac
2	1279	100.0	265	5 AAE26878	Aae26878 Helicobac
3	1270	99.3	253	2 AAW20486	Aaw20486 H. pylori
4	1270	99.3	253	2 AAW24673	Aaw24673 H. pylori
5	1270	99.3	256	2 AAW20866	Aaw20866 H. pylori
6	1264	98.8	253	5 AAE26880	Aae26880 Helicobac
7	1264	98.8	265	5 AAE26879	Aae26879 Helicobac
8	472.5	36.9	155	2 AAW98774	Aaw98774 H. pylori
9	252	19.7	49	5 AAE26876	Aae26876 Helicobac
10	198	15.5	38	5 AAE26873	Aae26873 Helicobac
11	194	15.2	41	5 AAE26875	Aae26875 Helicobac
12	152	11.9	30	5 AAE26874	Aae26874 Helicobac
13	150	11.7	30	5 AAE26872	Aae26872 Helicobac
14	99	7.7	222	4 AAB98720	Aab98720 Human bre
15	99	7.7	274	5 ABB40434	Abp40434 Staphyloc
16	98.5	7.7	1009	6 ABU15908	Abu15908 Protein e
17	98	7.7	913	6 ABM71523	Abm71523 Staphyloc
18	97	7.6	916	6 ABU48566	Abu48566 Protein e
19	96.5	7.5	701	5 ABP73563	Abp73563 Candida a
20	96	7.5	173	2 AAW44365	Aaw44365 Breast ca
21	96	7.5	173	4 AAB98717	Aab98717 Human bre
22	95	7.4	517	6 ABU26619	Abu26619 Protein e
23	95	7.4	2125	5 ABG77172	Abg77172 Prostate
24	95	7.4	2649	7 ABB79929	Abb79929 Human put
25	93.5	7.3	1196	6 ABU24813	Abu24813 Protein e

## ALIGNMENTS

## RESULT 1

AAE26860  
ID AAE26860 standard; protein; 253 AA.

XX AAE26860;

XX AC

XX 13-DEC-2002 (first entry)

XX DE Helicobacter sp. HP30 protein.

XX KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;  
low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
immunostimulant; cytostatic; peptic ulcer.

XX OS Helicobacter sp.

XX PN WO200251237-A2.

XX PD 04-JUL-2002.

XX PF 37-DEC-2001; 2001WO-US048392.

XX PR 37-DEC-2000; 2000US-00732091.

XX PA (ANTE-) ANTEX BIOLOGICS INC.

XX PI Tian J, Walker R, Jackson WJ;

XX DR WPI; 2002-666854/71.

XX N-PSDB; AAD44513.

XX PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
the proteins, useful as vaccines for raising immune response in animals.

XX Claim 1; Page 111-112; 127pp; English.

XX CC The invention relates to Helicobacter HP30 or HP56 polypeptide and  
peptides derived from them. The invention is useful for producing an  
immune response. It is useful for preventing, treating or ameliorating a  
disorder or disease associated with infection of an animal with  
Helicobacter. Pharmaceutical composition and vaccines comprising the  
sequences of the invention is useful for treating type B gastritis,  
peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B  
cell lymphoma. The invention is useful as reagents for clinical or  
medical diagnosis of Helicobacter infections and for scientific research  
on the properties of pathogenicity, virulence and infectivity of  
Helicobacter, as well as host defense mechanisms. The present sequence is

Aag82981 S. epider  
Abu13293 Protein e  
Aau33755 Staphyloc  
Aau36548 Staphyloc  
Abp30227 Streptoco  
Abp28162 Streptoco  
Aag44550 Arabidops  
Aag48726 Arabidops  
Aag44549 Arabidops  
Aag48725 Arabidops  
Aag48724 Arabidops  
Aag44548 Arabidops  
Aag30479 Streptoco  
Abp27884 Streptoco  
Abr52974 Protein s  
Aay51631 M. jannas  
Aay52002 M. jannas  
Aay31948 Plasmodiu  
Aay31947 Plasmodiu  
Aag82765 S. epider

26 93 7.3 381 4 AAG82981  
27 93 7.3 610 6 ABU13293  
28 92.5 7.2 396 4 AAU33755  
29 92.5 7.2 1009 4 AAU36548  
30 92 7.2 309 5 ABP30227  
31 92 7.2 327 5 ABP28162  
32 91.5 7.2 187 3 AAG44550  
33 91.5 7.2 187 3 AAG48726  
34 91.5 7.2 241 3 AAG44549  
35 91.5 7.2 241 3 AAG48725  
36 91.5 7.2 256 3 AAG48724  
37 91.5 7.2 256 3 AAG44548  
38 91.5 7.2 375 5 ABP30479  
39 91.5 7.2 390 5 ABP27884  
40 91.5 7.2 1769 6 ABR52974  
41 91.5 7.2 1847 3 AAY51631  
42 91.5 7.2 1847 3 AAY52002  
43 90.5 7.1 1411 2 AAY31948  
44 90.5 7.1 1417 2 AAY31947  
45 90 7.0 216 4 AAG82765

CC Helicobacter sp. HP30 protein  
 XX Sequence 253 AA;  
 SQ

Query Match 100.0%; Score 1279; DB 5; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-118;  
 Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYKYDRDLFLKQLESSDLDLFEVLVFGKGKGRHNEKLTSSIEYKRHGGDDYAKYAE 60  
 DB 1 MAYKYDRDLFLKQLESSDLDLFEVLVFGKGKGRHNEKLTSSIEYKRHGGDDYAKYAE 60

QY 61 IAEELQYVGSNSFASFIKGGVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
 DB 61 IAEELQYVGSNSFASFIKGGVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120

QY 121 LEEMDDREVKEMCDELSIKNTDNLNRQALSAAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180  
 DB 121 LEEMDDREVKEMCDELSIKNTDNLNRQALSAAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

QY 181 GRGLSLAGNOVLTRTSLFLGPGVGMIIIGVWTAIDIAGPAYRVTPACIIVATLRLKTOQ 240  
 DB 181 GRGLSLAGNOVLTRTSLFLGPGVGMIIIGVWTAIDIAGPAYRVTPACIIVATLRLKTOQ 240

QY 241 ANGDKKSQIESI 253  
 DB 241 ANGDKKSQIESI 253

RESULT 2  
 AA26878  
 ID AAE26878 standard; protein; 265 AA.  
 AC AAE26878;  
 XX  
 XX 13-DEC-2002 (first entry)  
 DT  
 DE Helicobacter sp. PQE/HP30 protein.  
 DE  
 KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
 KW type 8 gastritis; antiinflammatory; adenocarcinoma; defense mechanism;  
 KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
 KW immunostimulant; cytostatic; peptic ulcer.  
 XX  
 OS Helicobacter sp.  
 XX

Key Location/Qualifiers  
 FH Misc-difference 12 /note= "Encoded by TTC"  
 FT  
 FT  
 XX WC200251237-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 XX 07-DEC-2001; 2001WO-US048392.  
 XX  
 XX 07-DEC-2000; 2000US-00732091.  
 XX  
 XX (ANTE-) ANTEX BIOLOGICS INC.  
 PA  
 XX Tian J, Walker R, Jackson WJ;  
 XX  
 XX WPI; 2002-666854/71.  
 DR N-PSDB; AAD44535.  
 DR  
 XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
 the proteins, useful as vaccines for raising immune response in animals.  
 PT  
 PS Disclosure; Page 122-123; 127pp; English.  
 CC  
 CC The invention relates to Helicobacter HP30 or HP56 polypeptide and  
 CC peptides derived from them. The invention is useful for producing an  
 CC immune response. It is useful for preventing, treating or ameliorating a

CC disorder or disease associated with infection of an animal with  
 CC Helicobacter. Pharmaceutical composition and vaccines comprising the  
 CC sequences of the invention is useful for treating type B gastritis,  
 CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B  
 CC cell lymphoma. The invention is useful as reagents for clinical or  
 CC medical diagnosis of Helicobacter infections and for scientific research  
 CC on the properties of pathogenicity, virulence and infectivity of  
 CC Helicobacter, as well as host defense mechanisms. The present sequence is  
 CC Helicobacter sp. PQE/HP30 protein  
 XX  
 SQ Sequence 265 AA;  
 Query Match 100.0%; Score 1279; DB 5; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-118;  
 Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAYKYDRDLFLKQLESSDLDLFEVLVFGKGKGRHNEKLTSSIEYKRHGGDDYAKYAE 60  
 DB 13 MAYKYDRDLFLKQLESSDLDLFEVLVFGKGKGRHNEKLTSSIEYKRHGGDDYAKYAE 72

QY 61 IAEELQYVGSNSFASFIKGGVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
 DB 73 IAEELQYVGSNSFASFIKGGVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 132

QY 121 LEEMDDREVKEMCDELSIKNTDNLNRQALSAAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180  
 DB 133 LEEMDDREVKEMCDELSIKNTDNLNRQALSAAATLTLFKMGGFKSYQLAVIVANAVAKTIL 192

QY 181 GRGLSLAGNOVLTRTSLFLGPGVGMIIIGVWTAIDIAGPAYRVTPACIIVATLRLKTOQ 240  
 DB 193 GRGLSLAGNOVLTRTSLFLGPGVGMIIIGVWTAIDIAGPAYRVTPACIIVATLRLKTOQ 252

QY 241 ANGDKKSQIESI 253  
 DB 253 ANGDKKSQIESI 265

RESULT 3  
 AA20486  
 ID AAW20486 standard; protein; 253 AA.  
 AC AAW20486;  
 XX  
 XX 29-JUL-1997 (first entry)  
 DT  
 DE H. pylori cytoplasmic protein, 4095342.aa.  
 DE  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 XX  
 OS Helicobacter pylori.  
 XX  
 XX WC9640893-A1.  
 XX  
 XX 19-DEC-1996.  
 XX  
 XX 06-JUN-1996; 96WO-US009122.  
 XX  
 XX 07-JUN-1995; 95US-00487032.  
 PR 01-APR-1996; 96US-00630405.  
 XX  
 XX (ASTR) ASTRA AB.  
 PA  
 XX Smith D, Berglindh OT, Mellgaard BL;  
 PI WPI; 1997-052306/05.  
 DR N-PSDB; AAT67811.  
 DR  
 XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 PT useful for vaccines to treat or prevent H. pylori infection, and to  
 PT detect Helicobacter.  
 XX

PS Claim 6:-; Page 651; 1481pp; English.

XX The present sequence is a H. pylori cytoplasmic protein. The protein may  
 CC be used in a vaccine to prevent or treat H. pylori infection or to  
 CC identify H. pylori polypeptide binding compounds, useful as potential H.  
 CC pylori life cycle activators or inhibitors. The genomic sequence of H.  
 CC pylori (ATCC 55679) was determined from overlapping contigs generated by  
 CC mechanically shearing the bacterial DNA. The sequences were analysed for  
 CC ORF of at least 180 nucleotides, and the predicted coding regions defined  
 CC by computer evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF were  
 CC analysed for significant homology to other known or exported membrane  
 CC proteins. Having identified and determined the sequences of interest;  
 CC particular regions can be isolated from H. pylori by PCR amplification  
 CC for recombinant polypeptide production, e.g. in E. coli hosts

SQ Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;  
 Best Local Similarity 99.2%; Pred. No. 1.4e-117;  
 Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAYKYDRDLLEFLKQLESDDLDFEVLVFGKDGKRRHNEKLTSSIEYKRGHGGDYAKYAE 60  
 DB : MAYKYDRDLLEFLKQLESDDLDFEVLVFGKDGKRRHNEKLTSSIEYKRGHGGDYAKYAE 60

QY 61 IAEELQYSGNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
 DB 61 IAEELQYSGNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120

QY 121 LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTLFKMGGFKSYQLAVIVANAVAKTIL 180  
 DB 121 LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

QY 181 GRGLSLAGNOVLTRTSLFSGTGPVGMIIIGVWTAIDTAGPAYRVVTIPACIVVATLRKLTQQ 240  
 DB 181 GRGLSLAGNOVLTRTSLFSGTGPVGMIIIGVWTAIDTAGPAYRVVTIPACIVVATLRKLTQQ 240

QY 241 ANGDKKSLOIESI 253  
 DB 241 ANEDKKSLOIESW 253

RESULT 4  
 AAW24673  
 ID AAW24673 standard; protein; 253 AA.  
 AC AAW24673.  
 XX 12-AUG-1997 (first entry)

XX H. pylori cytoplasmic protein, 4095342.aa.  
 XX Transmembrane; cytoplasmic; cell envelope; flagella; transport; secreted;  
 XX periplasmic; chronic gastritis; duodenal ulcer disease; activator;  
 XX inhibitor; bacterial life cycle; vaccine; immunise; detection; actinase;  
 XX inhibition.

XX Helicobacter pylori.  
 XX WC9719098-A1.  
 XX 29-MAY-1997.  
 XX 15-NOV-1996; 96WC-US018542.  
 XX 17-NOV-1995; 95US-00561469.  
 XX (ASTR ) ASTRA AB.  
 XX Smith DH;  
 XX WPI; 1997-298052/27.

DR N-PSDB; AAT77491.

XX Helicobacter pylori nucleic acid sequences and related proteins - used  
 PT for diagnostics and therapeutics.

XX Claim 18; Page 184; 235pp; English.

XX This sequence represents an H. pylori cytoplasmic protein. Helicobacter  
 CC pylori has been strongly linked to chronic gastritis and duodenal ulcer  
 CC disease. The nucleic acid sequences of the invention are used to evaluate  
 CC compounds, especially activators or inhibitors of bacterial life cycle,  
 CC for the ability to bind an H. pylori nucleic acid sequence. The nucleic  
 CC acid sequences, and corresponding proteins, are also useful for  
 CC generating vaccines for immunising subjects against H. pylori or for use  
 CC in detecting the presence of Helicobacter species in a sample. Antisense  
 CC nucleic acid sequences of these sequences are used to inhibit expression  
 CC of a gene from Helicobacter species. H. pylori whole genomic DNA was  
 CC isolated and recombined to a median size of 2000 bp. Purified DNA  
 CC fragments were blunt-ended and ligated to unique BstXI-linker adapters in  
 CC 100-1000 fold molar excess. These linkers are complementary to the BstXI-  
 CC cut PMX vectors, while the overhang is not self-complementary. Therefore  
 CC the linkers will not concatamerise nor will the cut vector re-ligate  
 CC itself easily. The linker-adaptor inserts were ligated to each of the 20  
 CC PMX vectors to construct a series of shotgun subclone libraries. The  
 CC purified DNA samples were then sequenced. Note: The ORF/protein reference  
 CC number for this sequence was obtained from the related specification,  
 CC WO9640893

XX Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;  
 Best Local Similarity 99.2%; Pred. No. 1.4e-117;  
 Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAYKYDRDLLEFLKQLESDDLDFEVLVFGKDGKRRHNEKLTSSIEYKRGHGGDYAKYAE 60  
 DB 1 MAYKYDRDLLEFLKQLESDDLDFEVLVFGKDGKRRHNEKLTSSIEYKRGHGGDYAKYAE 60

QY 61 IAEELQYSGNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
 DB 61 IAEELQYSGNSFASFIKGEGLVYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120

QY 121 LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTLFKMGGFKSYQLAVIVANAVAKTIL 180  
 DB 121 LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

QY 181 GRGLSLAGNOVLTRTSLFSGTGPVGMIIIGVWTAIDTAGPAYRVVTIPACIVVATLRKLTQQ 240  
 DB 181 GRGLSLAGNOVLTRTSLFSGTGPVGMIIIGVWTAIDTAGPAYRVVTIPACIVVATLRKLTQQ 240

QY 241 ANGDKKSLOIESI 253  
 DB 241 ANEDKKSLOIESW 253

RESULT 5  
 AAW20866  
 ID AAW20866 standard; protein; 256 AA.  
 AC AAW20866;  
 XX 18-JUL-1997 (first entry)

XX H. pylori cytoplasmic protein, 12ge20305orf30.  
 XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 XX identification; binding compound; bacterium; life cycle; activator;  
 XX bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 XX diagnosis.  
 XX Helicobacter pylori.  
 XX WO9640893-A1.

```

XX 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US009122.
XX
XX 07-JUN-1995; 95US-00487032.
XX
XX 01-APR-1996; 96US-00630405.
XX
XX (ASTR ) ASTRA AB.
XX
XX Smith D, Berglindh OT, Mellgaerd BL;
XX
XX WPI; 1997-052306/05.
XX
XX N-PSDB; AAT68119.
XX
XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
XX useful for vaccines to treat or prevent H. pylori infection, and to
XX detect Helicobacter.
XX
XX Claim 61; Page 1269; 1481pp; English.
XX
XX This sequence represents a H. pylori cytoplasmic protein. The protein may
XX be used in a vaccine to prevent or treat H. pylori infection or to
XX identify H. pylori polypeptide binding compounds, useful as potential H.
XX pylori life cycle activators or inhibitors. The genomic sequence of H.
XX pylori (ATCC 55679) was determined from overlapping contigs generated by
XX mechanically shearing the bacterial DNA. The sequences were analysed for
XX ORF of at least 180 nucleotides, and the predicted coding regions defined
XX by computer evaluation. To identify likely H. pylori antigens for vaccine
XX development, the amino acid sequences predicted from various ORF were
XX analysed for significant homology to other known or exported membrane
XX proteins. Having identified and determined the sequences of interest,
XX particular regions can be isolated from H. pylori by PCR amplification
XX for recombinant polypeptide production, e.g. in E. coli hosts
XX
XX Sequence 256 AA;
XX
XX Query Match 99.3%; Score 1270; DB 2; Length 256;
XX Best Local Similarity 99.2%; Pred. No. 1.5e-117;
XX Matches 253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MAYKYDRDLBFLKQLESDDLDFEVLVFGKDGKRNHNEKLTSSIEYKRGHDDYAKYAE 60
XX DB 4 MAYKYDRDLBFLKQLESDDLDFEVLVFGKDGKRNHNEKLTSSIEYKRGHDDYAKYAE 63
XX
XX QY 61 IAEELQYVGSNSFASFIKGEVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120
XX DB 64 IAEELQYVGSNSFASFIKGEVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 123
XX
XX QY 121 LEEMDDDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFPMGFGKSYQLAVIVANAVAKTIL 180
XX DB 124 LEEMDDDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFPMGFGKSYQLAVIVANAVAKTIL 183
XX
XX QY 181 GRGLSLAGNOVLTRTSLFSLGPGVGMIIITGVWTAIDAGPAYRVTPACIVVATLRKTQQ 240
XX DB 184 GRGLSLAGNOVLTRTSLFSLGPGVGMIIITGVWTAIDAGPAYRVTPACIVVATLRKTQQ 243
XX
XX QY 241 ANGDKKSLQIESI 253
XX DB 244 ANEDKKSLQIESV 256
XX
XX RESULT 6
XX AAE26880
XX ID AAE26880 standard; protein; 253 AA.
XX
XX AC AAE26880;
XX
XX DT 13-DEC-2002 (first entry)
XX
XX Helicobacter sp. pOE/HP30 plasmid DNA encoded protein.
XX
XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;

```

```

KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytotatic; peptic ulcer.
XX
XX Helicobacter sp.
XX
XX WO200251237-A2.
XX
XX 04-JUL-2002.
XX
XX 07-DEC-2001; 2001WO-US048392.
XX
XX 07-DEC-2000; 2000US-00732091.
XX
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Tian J, Walker R, Jackson WJ;
XX
XX WPI; 2002-666854/71.
XX
XX N-PSDB; AAD44537.
XX
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
XX the proteins, useful as vaccines for raising immune response in animals.
XX
XX Disclosure; Page 126-127; 127pp; English.
XX
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and
XX peptides derived from them. The invention is useful for producing an
XX immune response. It is useful for preventing, treating or ameliorating a
XX disorder or disease associated with infection of an animal with
XX Helicobacter. Pharmaceutical composition and vaccines comprising the
XX sequences of the invention is useful for treating type B gastritis,
XX peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
XX cell lymphoma. The invention is useful as reagents for clinical or
XX medical diagnosis of Helicobacter infections and for scientific research
XX on the properties of pathogenicity, virulence and infectivity of
XX Helicobacter, as well as host defense mechanisms. The present sequence is
XX Helicobacter sp. pOE/HP30 plasmid DNA encoded protein
XX
XX Sequence 253 AA;

```

```

XX Query Match 98.8%; Score 1264; DB 5; Length 253;
XX Best Local Similarity 98.8%; Pred. No. 5.6e-117;
XX Matches 250; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MAYKYDRDLBFLKQLESDDLDFEVLVFGKDGKRNHNEKLTSSIEYKRGHDDYAKYAE 60
XX DB 1 MAYKYDRDLBFLKQLESDDLDFEVLVFGKDGKRNHNEKLTSSIEYKRGHDDYAKYAE 60
XX
XX QY 61 IAEELQYVGSNSFASFIKGEVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120
XX DB 61 IAEELQYVGSNSFASFIKGEVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120
XX
XX QY 121 LEEMDDDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFPMGFGKSYQLAVIVANAVAKTIL 180
XX DB 121 LEEMDDDEEVKEMCDELSIKNTDNLNRQALSAAATLTLFPMGFGKSYQLAVIVANAVAKTIL 180
XX
XX QY 181 GRGLSLAGNOVLTRTSLFSLGPGVGMIIITGVWTAIDAGPAYRVTPACIVVATLRKTQQ 240
XX DB 181 GRGLSLAGNOVLTRTSLFSLGPGVGMIIITGVWTAIDAGPAYRVTPACIVVATLRKTQQ 240
XX
XX QY 241 ANGDKKSLQIESI 253
XX DB 241 ANEDKKSLQIESI 253
XX
XX RESULT 7
XX AAE26879
XX ID AAE26879 standard; protein; 265 AA.
XX
XX AC AAE26879;
XX
XX DT 13-DEC-2002 (first entry)

```

```

XX Helicobacter sp. pQE/HP56 plasmid DNA encoded protein.
DE
XX
XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
XX type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
XX low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
XX immunostimulant; cystostatic; peptic ulcer.
XX
XX Helicobacter sp.
XX
XX WO200251237-A2.
XX
XX 04-JUL-2002.
XX
XX 07-DEC-2001; 2001WO-US048392.
XX
XX 07-DEC-2000; 2000US-00732091.
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Tian J, Walker R, Jackson WJ;
XX
XX WPI; 2002-669854/71.
XX N-PSDB; AAD44536.
XX
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
XX the proteins, useful as vaccines for raising immune response in animals.
XX
XX Disclosure; Page 124-125; 127pp; English.
XX
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and
XX peptides derived from them. The invention is useful for producing an
XX immune response. It is useful for preventing, treating or ameliorating a
XX disorder or disease associated with infection of an animal with
XX Helicobacter. Pharmaceutical composition and vaccines comprising the
XX sequences of the invention is useful for treating type B Gastritis,
XX peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
XX cell lymphoma. The invention is useful as reagents for clinical or
XX medical diagnosis of Helicobacter infections and for scientific research
XX on the properties of pathogenicity, virulence and infectivity of
XX Helicobacter, as well as host defense mechanisms. The present sequence is
XX Helicobacter sp. pQE/HP56 plasmid DNA encoded protein
XX
XX Sequence 265 AA;
XX
Query Match 98.8%; Score 1264; DB 5; Length 265;
Best Local Similarity 98.8%; Pred. No. 6e-117;
Matches 250; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MAYKYDRDLFLKQLESSLDDLFVLVFGKGEKRNKELTSSIEYKRGDDYAKYAE 60
DB 13 MAYKYDRDLFLKQLESSLDDLFVLVFGKGEKRNKELTSSIEYKRGDDYAKYAE 72
QY 61 IAEELQYGSNSPSPFIKGEGLVLYKEILCDVCKLVKNYKKTETTLIEQNMLSKILERS 120
DB 73 IAEELQYGSNSPSPFIKGEGLVLYKEILCDVCKLVKNYKKTETTLIEQNMLSKILERS 132
QY 121 LSEMDDEEVKEMCDELSIKNTQNTNROALSAAATLTLFKMGGFSYQLAVIVANAVAKTIL 180
DB 133 LSEMDDEEVKEMCDELSIKNTQNTNROALSAAATLTLFKMGGFSYQLAVIVANAVAKTIL 192
QY 181 GRGLSLAGQVLTFTLSFTLTPGVWIIITGVWTAIDTAGPAYRVTIIPACIVWATLRKLTQQ 240
DB 193 GRGLSLAGQVLTFTLSFTLTPGVWIIITGVWTAIDTAGPAYRVTIIPACIVWATLRKLTQQ 252
QY 241 ANGDKFSKLTQESI 253
DB 253 ANEDKFSKLTQESI 265
XX
RESULT 8
ID AAW98774 standard; protein; 155 AA.

```

```

XX AAW98774;
XX
XX 31-MAR-1999 (first entry)
XX
XX H. pylori GHPO 1170 protein.
XX
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX peptic ulcer disease.
XX
XX Helicobacter pylori.
XX
XX WO9843478-A1.
XX
XX 08-OCT-1998.
XX
XX 01-APR-1998; 98WO-US006371.
XX
XX 01-APR-1997; 97US-00833457.
XX
XX 24-JUN-1997; 97US-00881227.
XX
XX 29-JUL-1997; 97US-00902615.
XX
XX (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;
XX
XX WPI; 1998-542293/46.
XX N-PSDB; AAX14493.
XX
XX New isolated Helicobacter polynucleotides - used to develop products for
XX the diagnosis, prevention and treatment of Helicobacter infections and
XX gastrointestinal diseases.
XX
XX Claim 8; Page 1676-1677; 2054pp; English.
XX
XX This sequence represents a Helicobacter pylori GHPO protein of the
XX invention. The polypeptides can be used for preventing or treating
XX Helicobacter infections, and gastroduodenal diseases associated with
XX these infections, including acute, chronic, and atrophic gastritis, and
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
XX used for the production of antibodies. The products can also be used for
XX detection and diagnosis
XX
XX Sequence 155 AA;
XX
Query Match 36.9%; Score 472.5; DB 2; Length 155;
Best Local Similarity 63.3%; Pred. No. 1.3e-38;
Matches 93; Conservative 22; Mismatches 29; Indels 3; Gaps 1;
QY 38 NEKLTSSIEYKRGDDYAKYAEIABEELQYGSNSPSPFIKGEGLVLYKEILCDVCKLV 97
DB 2 NEKLTSSIEYKRGDDYAKYAEIABEELQYGSNSPSPFIKGEGLVLYKEILCDVCKLV 61
QY 98 NYNKKTTETTLIEQNMLSKILERSLEEMDDDEEVKEMCDELSIKNTDML---NRQALSATL 154
DB 62 NYNERSATSLIEQNMLSKILKDSLEKMSRREIKELCNELGTMIDKIVIGENKQVLIATSL 121
QY 155 TLFKMGGFSYQLAVIVANAVAKTILG 181
DB 122 TLFKAGGSHSYALAVSVADAMVRQTILG 148
XX
RESULT 9
AAE26876
ID AAE26876 standard; peptide; 49 AA.
XX
XX AAE26876;
XX
XX 13-DEC-2002 (first entry)
XX
XX Helicobacter sp. HP30-derived peptide #5.
XX

```

KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
 KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;  
 KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
 KW immunostimulant; cytostatic; peptic ulcer.  
 XX Helicobacter sp.  
 OS  
 XX WO200251237-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 07-DEC-2001; 2001WO-US048392.  
 XX  
 XX 07-DEC-2000; 2000US-00732091.  
 FR  
 XX (ANTE-) ANTEK BIOLOGICS INC.  
 XX  
 XX Tian J, Walker R, Jackson WJ;  
 XX  
 XX WPI; 2002-666854/71.  
 DR N-PSDB; AAD44529.  
 XX  
 XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
 PT the proteins, useful as vaccines for raising immune response in animals.  
 PT  
 XX Claim 7; Page 116; 127pp; English.  
 XX  
 CC The invention relates to Helicobacter HP30 or HP56 polypeptide and  
 CC peptides derived from them. The invention is useful for producing an  
 CC immune response. It is useful for preventing, treating or ameliorating a  
 CC disorder or disease associated with infection of an animal with  
 CC Helicobacter. Pharmaceutical composition and vaccines comprising the  
 CC sequences of the invention is useful for treating type B gastritis,  
 CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B  
 CC cell lymphoma. The invention is useful as reagents for clinical or  
 CC medical diagnosis of Helicobacter infections and for scientific research  
 CC on the properties of pathogenicity, virulence and infectivity of  
 CC Helicobacter, as well as host defense mechanisms. The present sequence is  
 CC Helicobacter sp. HP30-derived peptide  
 XX  
 SQ Sequence 49 AA;  
 Query Match 19.7%; Score 252; DB 5; Length 49;  
 Best Local Similarity 100.0%; Pred. No. 2e-17;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 203 VGMITGVWTAIDAGPAYRVTIPACIVWATRLKLTQQANGDKKSLQIE 251  
 DB 1 VGMITGVWTAIDAGPAYRVTIPACIVWATRLKLTQQANGDKKSLQIE 49  
 RESULT 10  
 AAE26873  
 ID AAE26873 standard; peptide; 38 AA.  
 AC  
 AC AAE26873;  
 XX  
 XX 13-DEC-2002 (first entry)  
 DE Helicobacter sp. HP30-derived peptide #2.  
 XX  
 XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
 KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;  
 KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
 KW immunostimulant; cytostatic; peptic ulcer.  
 XX Helicobacter sp.  
 OS  
 XX WO200251237-A2.  
 PN  
 XX 04-JUL-2002.  
 PD  
 XX 07-DEC-2001; 2001WO-US048392.  
 PF

XX 07-DEC-2000; 2000US-00732091.  
 XX (ANTE-) ANTEK BIOLOGICS INC.  
 XX  
 XX Tian J, Walker R, Jackson WJ;  
 XX  
 XX WPI; 2002-666854/71.  
 DR N-PSDB; AAD44526.  
 XX  
 XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
 PT the proteins, useful as vaccines for raising immune response in animals.  
 PT  
 XX Claim 7; Page 115; 127pp; English.  
 XX  
 CC The invention relates to Helicobacter HP30 or HP56 polypeptide and  
 CC peptides derived from them. The invention is useful for producing an  
 CC immune response. It is useful for preventing, treating or ameliorating a  
 CC disorder or disease associated with infection of an animal with  
 CC Helicobacter. Pharmaceutical composition and vaccines comprising the  
 CC sequences of the invention is useful for treating type B gastritis,  
 CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B  
 CC cell lymphoma. The invention is useful as reagents for clinical or  
 CC medical diagnosis of Helicobacter infections and for scientific research  
 CC on the properties of pathogenicity, virulence and infectivity of  
 CC Helicobacter, as well as host defense mechanisms. The present sequence is  
 CC Helicobacter sp. HP30-derived peptide  
 XX  
 SQ Sequence 38 AA;  
 Query Match 15.5%; Score 198; DB 5; Length 38;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-12;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 53 DYAKYAEIAEELQYVGSNSFASFIKGEVLYKEILCD 90  
 DB 1 DYAKYAEIAEELQYVGSNSFASFIKGEVLYKEILCD 38  
 RESULT 11  
 AAE26875  
 ID AAE26875 standard; peptide; 41 AA.  
 AC  
 AC AAE26875;  
 XX  
 XX 13-DEC-2002 (first entry)  
 DE Helicobacter sp. HP30-derived peptide #4.  
 XX  
 XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
 KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;  
 KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
 KW immunostimulant; cytostatic; peptic ulcer.  
 XX Helicobacter sp.  
 OS  
 XX WO200251237-A2.  
 PN  
 XX 04-JUL-2002.  
 PD  
 XX 07-DEC-2001; 2001WO-US048392.  
 XX  
 XX 07-DEC-2000; 2000US-00732091.  
 XX (ANTE-) ANTEK BIOLOGICS INC.  
 XX  
 XX Tian J, Walker R, Jackson WJ;  
 XX  
 XX WPI; 2002-666854/71.  
 DR N-PSDB; AAD44528.  
 XX  
 XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
 PT the proteins, useful as vaccines for raising immune response in animals.  
 PT









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:10:43 ; Search time 14 Seconds  
(without alignments)  
1738.318 Million cell updates/sec

Title: JS-09-732-091-4  
Perfect score: 1279  
Sequence: 1 MAYKYDRDLEFLKQLESSDL.....LRLKTOQANGDKKSQIESI 253

Scoring table: BLOSUM62  
Gapcp 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1279	100.0	253	D64718	conserved hypothet
2	1270	99.3	253	B71800	hypothetical prote
3	784.5	61.3	207	E64718	conserved hypothet
4	630	49.3	209	A71800	hypothetical prote
5	472.5	36.9	155	C64718	conserved hypothet
6	265	20.7	237	C64721	hypothetical prote
7	265	20.7	237	AC0503	conserved hypothet
8	262	20.5	237	C90630	probable oxidoredu
9	262	20.5	237	C85481	probable oxidoredu
10	133	10.4	39	F64718	hypothetical prote
11	98.5	7.7	1009	C89910	hypothetical prote
12	98	7.7	282	AE0360	hypothetical prote
13	98	7.7	1039	G90546	conserved hypothet
14	97.5	7.6	311	C81380	probable D-2-Hydro
15	97.5	7.6	728	A81385	probable ATP /GTP
16	97	7.6	916	E71330	probable preprotel
17	96.5	7.5	373	G70355	8-amino-7-oxononan
18	96	7.5	262	T33071	hypothetical prote
19	96	7.5	420	F96920	probable permeal
20	95.5	7.5	344	A47025	abortive phage res
21	95	7.4	517	C81327	hypothetical prote
22	95	7.4	2649	A40937	bullicous pemphigoid
23	93.5	7.3	802	S49252	penicillin amidase
24	93	7.3	355	E90488	transposase ISC122
25	93	7.3	610	G70178	exodeoxyribonuclea
26	92.5	7.2	430	T33070	hypothetical prote
27	92.5	7.2	802	I39665	penicillin amidase
28	92.5	7.2	886	T16536	hypothetical prote
29	92	7.2	308	H81345	hypothetical prote

30 92 7.2 1939 2 T18372 repeat organellar  
31 91.5 7.2 399 2 G72253 RNA polymerase sig  
32 91.5 7.2 472 2 A64320 PRT12 homolog - M  
33 91.5 7.2 508 2 T24622 hypothetical prote  
34 91.5 7.2 615 2 S42797 rad 26 protein - f  
35 91.5 7.2 1769 2 S53378 probable membrane  
36 91.5 7.2 1847 2 E64477 replication factor  
37 91 7.1 284 2 A45488 body-wall muscle t  
38 91 7.1 359 2 A12368 hypothetical prote  
39 90.5 7.1 320 2 B97214 hypothetical prote  
40 90.5 7.1 1017 2 D90550 vsaa-like (mycopla  
41 90.5 7.1 1411 2 T18417 hypothetical prote  
42 90.5 7.1 1417 2 T18418 hypothetical prote  
43 90 7.0 404 1 S03849 ribonucleoprotein  
44 89.5 7.0 280 2 A72046 conserved hypothet  
45 89.5 7.0 280 2 F86578 Cte71 hypothetical

## ALIGNMENTS

## RESULT 1

D64718  
conserved hypothetical protein HP1588 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: D64718  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:952185  
A:Accession: D64718  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-253 <TOM>  
A:Cross-references: GB:AE000656; GB:AE000511; NID:G2314771; PID:RAD08627.1; PID:G2314

Query Match 100.0%; Score 1279; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 4e-88;  
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAYKYDRDLEFLKQLESSDLDLFEVLVFGKGEKHNEKLTSSIEYKRGDDYAKYAE 60  
DB 1 MAYKYDRDLEFLKQLESSDLDLFEVLVFGKGEKHNEKLTSSIEYKRGDDYAKYAE 60  
QY 61 IAEELQYVGSNPFASFIKGEVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
DB 61 IAEELQYVGSNPFASFIKGEVLYKEILCDVCDKLVNKNKKTETTLIEQNMLSKILERS 120  
QY 121 LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTIFKMGGFKSYQLAVIVANAVAKTIL 180  
DB 121 LEEMDDEEVKEMCDLSIKNTDNLNRQALSATLTIFKMGGFKSYQLAVIVANAVAKTIL 180  
QY 181 GRGLSLAGNQVTRTSLZLTGPGVMTITGVTATDAIDAGPAYRVVTIPACIVATLRLKTOQ 240  
DB 181 GRGLSLAGNQVTRTSLZLTGPGVMTITGVTATDAIDAGPAYRVVTIPACIVATLRLKTOQ 240  
QY 241 ANGDKKSQIESI 253  
DB 241 ANGDKKSQIESI 253

## RESULT 2

B71800  
hypothetical protein jhp1494 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C:Accession: B71800  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*  
 A:Reference number: A71800; MUID:99120557; PMID:9923682  
 A:Accession: 371800  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-253 <ARN>  
 A:Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AAD07073.1; PID:g415612  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: jhp1494

Query Match 99.3%; Score 1270; DB 2; Length 253;  
 Best Local Similarity 99.2%; Pred. No. 1.9e-87;  
 Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MAYKYDRDLFLQLESSDLDLFEVLVFGKQKRNHKLTSIEYKRGHDDYAKYAE 60  
 DB 1 MAYKYDRDLFLQLESSDLDLFEVLVFGKQKRNHKLTSIEYKRGHDDYAKYAE 60  
 QY 61 IAEELQYVGNSSPASFKIGEGVLYKEILCDVCDKLKVNKKTETTLIEQNMLSKILERS 120  
 DB 61 IAEELQYVGNSSPASFKIGEGVLYKEILCDVCDKLKVNKKTETTLIEQNMLSKILERS 120  
 QY 121 LEEMDDEVKEMCDELSIKNTDNLNQAALSAATLTFKMGFKSYQLAVIVANAVAKTIL 180  
 DB 121 LEEMDDEVKEMCDELSIKNTDNLNQAALSAATLTFKMGFKSYQLAVIVANAVAKTIL 180  
 QY 181 GRGLSLAGNOVLTSLFPGVGIITGWTAIDAGPVRVTPACIVVATLRK 240  
 DB 181 GRGLSLAGNOVLTSLFPGVGIITGWTAIDAGPVRVTPACIVVATLRK 240  
 QY 241 ANGDKKSQIESI 253  
 DB 241 ANEDKKSQIESV 253

RESULT 3  
 E64718  
 conserved hypothetical protein HP1589 - *Helicobacter pylori* (strain 26695)  
 C:Species: *Helicobacter pylori*  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Aug-1997  
 C:Accession: E64718  
 R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, N. Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: E64718  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-207 <TOM>  
 A:Cross-references: GB:AE000511; TIGR:HP1589  
 C:Genetics:  
 A:Start codon: TTG

Query Match 61.3%; Score 784.5; DB 2; Length 207;  
 Best Local Similarity 73.6%; Pred. No. 2.2e-51;  
 Matches 153; Conservative 21; Mismatches 33; Indels 1; Gaps 1;  
 QY 41 LTSIEYKRGHDDYAKYAEIAEELQYVGNSSPASFKIGEGVLYKEILCDVCDKLKVN 100  
 DB 1 MTSIEYORYGYDYAKYPRRIAEELQYVGNSSPASFKIGEGVLYKEILCDVCDKLKVN 60  
 QY 101 KYTETTLIEQNMLSKILERSLEEMDDEVKEMCDELSIKNTDNLNQAALSAATLTFKMG 160  
 DB 61 KXSPTTIEENMUSSIIQKSLKMSDEIRLCELDGKVNKLGKQALSTAAATLTFKMG 120  
 QY 161 GFKSYQLAVIVANAVAKTILGRGLSLAGNOVLTSLFPGVGIITGWTAIDAGP 220

Db 121 GFKSYQLAVIVANAVAKTILGRGLSLAGNOVLTSLFPGVGIITGWTAIDAGP 180  
 QY 221 YRVTPACIVVATLRKLTQOANGDKKSL 248  
 Db 181 YRVTPACIVVATLRKKA-QANBIKNIL 207  
 RESULT 4  
 A71800  
 hypothetical protein jhp1493 - *Helicobacter pylori* (strain J99)  
 C:Species: *Helicobacter pylori*  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
 C:Accession: A71800  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F. Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*  
 A:Reference number: A71800; MUID:99120557; PMID:9923682  
 A:Accession: A71800  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-209 <ARN>  
 A:Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AAD07072.1; PID:g415612  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: jhp1493

Query Match 49.3%; Score 630; DB 2; Length 209;  
 Best Local Similarity 59.8%; Pred. No. 7.4e-40;  
 Matches 122; Conservative 32; Mismatches 46; Indels 4; Gaps 2;  
 QY 38 NEKLTSSIEYKRGHDDYAKYAEIAEELQYVGNSSPASFKIGEGVLYKEILCDVCDKLK 97  
 DB 2 NEELTSLTEYORYGYDYAKYPRRIAEELQYVGNSSPASFKIGEGVLYKEILCDVCDKLK 61  
 QY 98 NYNKKTTTLLIEQNMLSKILERSLEEMDDEVKEMCDELSIKNTDNLNQAALSAATL 154  
 DB 62 NYNRSATSLIEQNMLSKILERSLEEMDDEVKEMCDELSIKNTDNLNQAALSAATL 121  
 QY 155 TLFKMGFKSYQLAVIVANAVAKTILGRGLSLAGNOVLTSLFPGVGIITGWTA 213  
 DB 122 TLFKMGFKSHSYALAVADAVMRVQTLGHLSSVVGKVALKKTLDILAGPVGWITGALVS 181  
 QY 214 IDIAGPVRVTPACIVVATLRK 237  
 DB 182 INLAGPVRVTPACIVVATLRK 205

RESULT 5  
 C64718  
 conserved hypothetical protein HP1587 - *Helicobacter pylori* (strain 26695)  
 C:Species: *Helicobacter pylori*  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 C:Accession: C64718  
 R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, N. Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: C64718  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-155 <TOM>  
 A:Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08626.1; PID:g2314771

Query Match 36.9%; Score 472.5; DB 2; Length 155;  
 Best Local Similarity 63.3%; Pred. No. 2.8e-28;  
 Matches 93; Conservative 22; Mismatches 29; Indels 3; Gaps 1;  
 QY 38 NEKLTSSIEYKRGHDDYAKYAEIAEELQYVGNSSPASFKIGEGVLYKEILCDVCDKLK 97

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 5, 2004, 01:15:47 ; Search time 64 Seconds  
(without alignments)  
6701.674 Million cell updates/sec

Title: US-09-732-091-3  
Perfect score: 1349  
Sequence: 1 atggcatcaaatatgatg.....cgttgcaaatagatccatt

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%  
Maximum Watch 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+n2p.model -DEV=xlh  
-O=/cp2 1/uspt0 spool/US09732091/runat 02072004 182406 18996/app query.fasta\_1.903  
-DB=A\_Geneseq 25Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09732091@cgn 1 101 @runat 02072004 182406 18996 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCGES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=2 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq 29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1279	94.8	253	5	AAE26860	Helicobac
2	1279	94.8	265	5	AAE26878	Helicobac
3	1270	94.1	253	2	AAW20486	H. pylori
4	1270	94.1	253	2	AAW24673	H. pylori
5	1270	94.1	256	2	AAW20866	H. pylori
6	1264	93.7	253	5	AAE26880	Helicobac
7	1264	93.7	265	5	AAE26879	Helicobac
8	472.5	35.0	155	2	AAW98774	H. pylori
9	252	18.7	49	5	AAE26876	Helicobac
10	198	14.7	38	5	AAE26873	Helicobac

11	194	14.4	41	5	AAE26875	Helicobac
12	152	11.3	30	5	AAE26874	Helicobac
13	150	11.1	30	5	AAE26872	Helicobac
14	101.5	7.5	173	4	AAW44365	Breast ca
15	101.5	7.5	173	4	AAW98717	Human bre
16	99	7.3	222	4	AAW98720	Human bre
17	99	7.3	274	5	ABP40434	Staphyloc
18	98.5	7.3	1009	6	ABU15908	Protein e
19	98	7.3	413	6	ABU1523	Staphyloc
20	97	7.2	916	6	ABU1523	Staphyloc
21	96.5	7.2	414	7	ABU1523	Staphyloc
22	96.5	7.2	701	5	ABU1523	Staphyloc
23	95	7.0	517	6	ABU1523	Staphyloc
24	95	7.0	2125	5	ABU1523	Staphyloc
25	95	7.0	2449	7	ABU1523	Staphyloc
26	93.5	6.9	1196	6	ABU1523	Staphyloc
27	93	6.9	381	4	ABU1523	Staphyloc
28	93	6.9	610	6	ABU1523	Staphyloc
29	93	6.9	667	2	AAW94919	Rat phero
30	92.5	6.9	996	4	AAU33755	Staphyloc
31	92.5	6.9	1009	4	AAU36548	Staphyloc
32	92	6.8	309	5	ABP30227	Streptoco
33	92	6.8	327	5	ABP30227	Streptoco
34	91.5	6.8	187	3	AAW44550	Arabisdops
35	91.5	6.8	187	3	AAW44550	Arabisdops
36	91.5	6.8	241	3	AAW44549	Arabisdops
37	91.5	6.8	241	3	AAW44549	Arabisdops
38	91.5	6.8	256	3	AAW44548	Arabisdops
39	91.5	6.8	375	5	ABP30479	Streptoco
40	91.5	6.8	375	5	ABP30479	Streptoco
41	91.5	6.8	390	5	ABP27884	Streptoco
42	91.5	6.8	1769	6	ABR52974	Protein s
43	91.5	6.8	1847	3	AAV51631	M. jannas
44	91.5	6.8	1847	3	AAV52002	M. jannas
45	90.5	6.7	1411	2	AAV31948	Plasmodiu

ALIGNMENTS

RESULT 1  
AAE26860  
ID AAE26860 standard; protein; 253 AA.

XX AC AAE26860;  
XX DT 13-DEC-2002 (first entry)  
XX DE Helicobacter sp. HP30 protein.  
XX KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
XX KW type B gastritis; anti-inflammatory; adenocarcinoma; defense mechanism;  
XX KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
XX KW immunostimulant; cytostatic; peptic ulcer.  
XX OS Helicobacter sp.  
XX PN W0200251237-A2.  
XX PD 04-JUL-2002.  
XX PF 07-DEC-2001; 2001WO-US048392.  
XX PR 07-DEC-2000; 2000US-00732091.  
XX (ANTE-) ANTEX BIOLOGICS INC.  
XX Tian J, Walker R, Jackson WJ;  
XX WPI: 2002-666854/71.  
XX N-PSDB: AAD44513.  
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
XX the proteins, useful as vaccines for raising immune response in animals.

XX Claim 1; Page 111-112; 127pp; English.

XX The invention relates to Helicobacter HP30 or HP56 polypeptide and

CC peptides derived from them. The invention is useful for producing an

CC immune response. It is useful for preventing, treating or ameliorating a

CC disorder or disease associated with infection of an animal with

CC Helicobacter. Pharmaceutical composition and vaccines comprising the

CC sequences of the invention is useful for treating type B gastritis,

CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B

CC cell lymphoma. The invention is useful as reagents for clinical or

CC medical diagnosis of Helicobacter infections and for scientific research

CC on the properties of pathogenicity, virulence and infectivity of

CC Helicobacter, as well as host defense mechanisms. The present sequence is

CC Helicobacter sp. HP30 protein

XX Sequence 253 AA;

SQ

Alignment Scores:

Pred. No.: 6.71e-135 Length: 253

Score: 1279.00 Matches: 253

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 94.81% Indels: 0

DB: 5 Gaps: 0

US-09-732-091-3 (1-759) x AAE26860 (1-253)

QY 1 ATGCATACAAATGATGAGACCTGCAATTTTAAAGCAATGGATCTAGTGATTTA 60

DB 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

QY 61 TTGATTGTTTGGAGTCTTTGTTTGGTAAAGACGCGGAAAGACACATCAAAA 120

DB 21 LeuAspLeuPheGluValLeuValPheGlyAspGlyGluLysArgHisAsnGluLys 40

QY 121 CTGACCGCTCCATAGATACAAAGGATCGGATGATTAAGCTTAATAGCAGAAAGA 180

DB 41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

QY 181 ATCCCTGAAGAGTTGCATACATCGGAGCAATGTTTGGAGTTTCCATTAAAGCGCAA 240

DB 61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

QY 241 GGAGCTCTTATACAAAGAGATTTTATGCGATGTGTGCGATTAATTAAGTCAATTACAC 300

DB 81 GlyValLeuTyrLysGluLeuLysAspValCysAspLysLeuLysValAsnTyrAsn 100

QY 301 AAGAACTGAACGACTTAAATGACAAACATGCTTTCTAAATCTTAGAAGAGT 360

DB 101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

QY 361 TTGAAGAATGGATGATGAAGAAGTGAAGAATGTCGATGAATTCATTAATCCATAAAAAC 420

DB 121 LeuGluGluMetAspAspGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

QY 421 ACAGCAATTTAAACAGACAGCTTAAGCGCGGCACTTTAACCTCTTTAAATGGGG 480

DB 141 ThrAspAsnLeuAspArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

QY 481 GGTTTTAATCTTATCAATAGCTCATGTTGGGAATGCGGTGCGCAAAACCACTTCTA 540

DB 161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

QY 541 GGGCTGTTTATCCCTGGCGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTAAACA 600

DB 181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

QY 601 GGTCTGTGGCTGATCATTAACGCGCTATGACAGCGATGATATTGCGGGCGCGCT 660

DB 201 GlyProValGlyTyrIleIleThrGlyValThrAlaIleAspIleAlaGlyProAla 220

QY 661 TATAGGTAACCATACCGGATGCAATGTTGGTTCACACTTTACGCTTAAACACAGCAA 720

DB 221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGln 240

QY 721 GCCAATCGCATAGAAAGTCTGTCAAATAGAACATCCATT 759

DB 241 AlaAsnGlyAspLysSerLeuGlnIleGluSerIle 253

RESULT 2

AAE26878

ID AAE26878 standard; protein; 265 AA.

XX AAE26878;

AC AAE26878;

XX 13-DEC-2002 (first entry)

DE Helicobacter sp. PQE/HP30 protein.

XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;

KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;

KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;

XX immunostimulant; cytostatic; peptic ulcer.

OS Helicobacter sp.

XX Key Location/Qualifiers

FT Misc-difference 12 /note= "Encoded by TTC"

XX WO200251237-A2.

XX 04-JUL-2002.

XX 07-DEC-2001; 2001WO-US048392.

XX 07-DEC-2000; 2000US-00732091.

XX (ANTE-) ANTEX BIOLOGICS INC.

PI Tian J, Walker R, Jackson WJ;

XX WPI; 2002-666854/71.

DR N-PSDB; AAD44535.

XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding

PT the proteins, useful as vaccines for raising immune response in animals.

XX Disclosure; Page 122-123; 127pp; English.

XX The invention relates to Helicobacter HP30 or HP56 polypeptide and

CC peptides derived from them. The invention is useful for producing an

CC immune response. It is useful for preventing, treating or ameliorating a

CC disorder or disease associated with infection of an animal with

CC Helicobacter. Pharmaceutical composition and vaccines comprising the

CC sequences of the invention is useful for treating type B gastritis,

CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B

CC cell lymphoma. The invention is useful as reagents for clinical or

CC medical diagnosis of Helicobacter infections and for scientific research

CC on the properties of pathogenicity, virulence and infectivity of

CC Helicobacter, as well as host defense mechanisms. The present sequence is

CC Helicobacter sp. PQE/HP30 protein

XX Sequence 265 AA;

SQ

Alignment Scores:

Pred. No.: 6.84e-135 Length: 265

Score: 1279.00 Matches: 253

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 94.81% Indels: 0

DB: 5 Gaps: 0

US-09-732-091-3 (1-759) x AAE26878 (1-265)

QY 1 ATGGCATCAAAATATAGACATTCGAATTTTAAAGCAATTCGAATCTAGTACTTA 60  
 Db 13 MetAlaTyLysTyAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 32  
 QY 61 TTGGATTGTTTACGCTGCTTTTGGTAAAGCGCGGAAAAAGACACAAATCAAAA 120  
 Db 33 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluYsArgHisGlnGluLys 52  
 QY 121 CTGACCACTCCATAGATACAAAAGCGATGCGATGATTAAGCTTAATATACGAGAAGA 180  
 Db 53 LeuThrSerSerIleGluTyLysArgHisGlyAspAspTyLysAlaLysTyAlaGluArg 72  
 QY 181 ATCGCTGAAGAGTTCATATACGACATAGTTTTCGAGTTTTCATTAAGCGGAA 240  
 Db 73 LeuAlaGluGluLeuGluTyLysSerAsnSerPheAlaSerPheIleLysGlyGlu 92  
 QY 241 GGAGCTCTTATACAAAGAGATTTTATCGCATGCTGCGATATAAATTAAGGTCAATTACAAC 300  
 Db 93 GlyValLeuTyLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyAsn 112  
 QY 301 AAGAAACTGAAACGACTTAATGAAACAAACATGCTTCTAAATCTTAGAAGAGT 360  
 Db 113 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 132  
 QY 361 TTGGAGAAATGATGATCAAGAGTGAAGAAATGCGGATGAATATCCATAAAAC 420  
 Db 133 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 152  
 QY 421 ACGGCAATTTAAACAGACAAAGCTTTAAGCGGCGGACTTTAACGCTGTTTAAATGGG 480  
 Db 153 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 172  
 QY 481 GCTTTAAATCTATCAATAGTGTCAATGTTGGAATGCGGATGCGCAAAACCATTTCTA 540  
 Db 173 GlyPheLysSerTyGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 192  
 QY 541 GGGCGTGTATTCGTTCCGGGCATCAGGTGCTTACAGACTCTGAGCTTTTAAACA 600  
 Db 193 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 212  
 QY 601 GTCCTGTGTTGGTGGATCATTACAGGCGATGACAGCGATTCATATTCAGGCGCGCT 660  
 Db 213 GlyProValGlyTyLysIleThrGlyValTyrThrAlaIleAspIleAlaGlyProAla 232  
 QY 661 TATAGGTAAACATACCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 Db 233 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 252  
 QY 721 GCCAATGGAGATGAAGAGTCGTTGCAATAGATCCATT 759  
 Db 253 A-AsnGlyAspLysLysSerLeuGlnIleGluSerIle 265

RESULT 3

AAW20486

ID AAW20486 standard; protein; 253 AA.

XX

AC AAW20486

XX

DT 29-JUL-1997 (first entry)

XX

DE H. pylori cytoplasmic protein, 4095342.aa.

XX

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

XX

OS Helicobacter pylori.

XX

PN W09640893-A1.

XX

PC 19-DEC-1996.

XX

PF 06-JUN-1996; 36WC-US009122.

XX 07-JUN-1995; 35US-00487032.  
 PR 01-APR-1996; 96US-00630405.  
 XX  
 XX (ASTR } ASTRA AB.  
 XX Smith D, Berglindh OT, Mellgaard BL;  
 PI WPI; 1997-052306/05.  
 XX N-PSDB; AAT67811.  
 DR  
 XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 PT useful for vaccines to treat or prevent H. pylori infection, and to  
 PT detect Helicobacter.  
 XX  
 PS Claim 61; Page 651; 1481pp; English.

XX The present sequence is a H. pylori cytoplasmic protein. The protein may  
 CC be used in a vaccine to prevent or treat H. pylori infection or to  
 CC identify H. pylori polypeptide binding compounds, useful as potential H.  
 CC pylori life cycle activators or inhibitors. The genomic sequence of H.  
 CC pylori (ATCC 55679) was determined from overlapping contigs generated by  
 CC mechanically shearing the bacterial DNA. The sequences were analysed for  
 CC ORF of at least 180 nucleotides, and the predicted coding regions defined  
 CC by computer evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF were  
 CC analysed for significant homology to other known or exported membrane  
 CC proteins. Having identified and determined the sequences of interest,  
 CC particular regions can be isolated from H. pylori by PCR amplification  
 CC for recombinant polypeptide production, e.g. in E. coli hosts

XX Sequence 253 AA;

Alignment Scores:

Pred. No.: 6,94e-134

Score: 1270.00

Percent Similarity: 99.60%

Best Local Similarity: 99.21%

Query Match: 94.14%

DB: 2

Length: 253

Matches: 1

Conservative: 1

Mismatches: 1

Indels: 0

Gaps: 0

US-09-732-091-3 (1-759) x AAW20486 (1-253)

QY 1 ATGGCATCAAAATATAGACATTCGAATTTTAAAGCAATTCGAATCTAGTACTTA 60  
 Db 1 MetAlaTyLysTyAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20  
 QY 61 TTGGATTGTTTACGCTGCTTTTGGTAAAGCGCGGAAAAAGACACAAATCAAAA 120  
 Db 21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisGlnGluLys 40  
 QY 121 CTGACCACTCCATAGATACAAAAGCGATGCGGATTAATACGCAAGAAAGA 180  
 Db 41 LeuThrSerSerIleGluTyLysArgHisGlyAspAspTyLysAlaLysTyAlaGluArg 60  
 QY 181 ATCGCTGAAGAGTTCATATACGACATAGTTTTCGAGTTTTCATTAAGCGGAA 240  
 Db 61 IleAlaGluGluLeuGluTyLysSerAsnSerPheAlaSerPheIleLysGlyGlu 80  
 QY 241 GGAGCTCTTATACAAAGAGATTTTATCGCATGCTGCGATATAAATTAAGGTCAATTACAAC 300  
 Db 81 GlyValLeuTyLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyAsn 100  
 QY 301 AAGAAACTGAAACGACTTAATGAAACAAACATGCTTCTAAATCTTAGAAGAGAGT 360  
 Db 101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120  
 QY 361 TTGGAGAAATGATGATGAAGAGTGAAGAAATGCGGATGAATATCCATAAAAC 420  
 Db 121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140  
 QY 421 ACGGCAATTTAAACAGACAAAGCTTTAAGCGGCGGACTTTAACGCTGTTTAAATGGG 480



Db 141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrIleuThrLeuPheLysMetGly 160  
 QY 481 GGTTTAAATCTTATCAATGCTGTCATTGTTGGCAATGCGTCCGCAAAACCATCTCTA 540  
 Db 161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180  
 QY 541 GGCGCTGGTTATPCGCTTGGCGCAATCAGGTGCTTACAGAACTCTGAGCTTTTAAACA 600  
 Db 181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValIleuThrArgThrLeuSerPheLeuThr 200  
 QY 601 GGTCTCTGCTGGTGGATCATATACAGCGGTATGACACGCGATTGATGTCAGGCGCGCT 660  
 Db 201 GlyProValGlyTyrPheIleThrGlyValThrAlaIleAspIleAlaGlyProAla 220  
 QY 661 TATAGGGAACCATACCGCATCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 720  
 Db 221 TyrArgValThrIleProAlaCysIleValAlaValAlaThrLeuArgLeuLysThrGlnGln 240  
 QY 721 GCCAATGGAGATAAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 759  
 Db 241 AlaAsnGluAspLysSerLeuGlnIleGluSerVal 253

## RESULT 4

AAW24673  
 ID AAW24673 standard; protein; 253 AA.

AC AAW24673

DT 12-AUG-1997 (first entry)

DE H. pylori cytoplasmic protein, 4095342.aa.

XX Transmembrane; cytoplasmic; cell envelope; flagella; transport; secreted;  
 KW periplasmic; chronic gastritis; duodenal ulcer disease; activator;  
 KW inhibitor; bacterial life cycle; vaccine; immunise; detection; antisense;  
 KW inhibition.

XX Helicobacter pylori.

XX W09719098-A1.

PD 29-MAY-1997.

XX 15-NOV-1996; 96WO-US018542.

XX 17-NOV-1995; 95JS-00561469.

XX (ASTR) ASTRA AB.

XX Smith DH;

XX WPI; 1997-298052/27.

XX N-PSDB; AAT77491.

XX Helicobacter pylori nucleic acid sequences and related proteins - used  
 PT for diagnostics and therapeutics.

XX Claim 18; Page 184; 235pp; English.

XX This sequence represents an H. pylori cytoplasmic protein. Helicobacter  
 CC pylori has been strongly linked to chronic gastritis and duodenal ulcer  
 CC disease. The nucleic acid sequences of the invention are used to evaluate  
 CC compounds, especially activators or inhibitors of bacterial life cycle,  
 CC for the ability to bind an H. pylori nucleic acid sequence. The nucleic  
 CC acid sequences, and corresponding proteins, are also useful for  
 CC generating vaccines for immunising subjects against H. pylori or for use  
 CC in detecting the presence of Helicobacter species in a sample. Antisense  
 CC nucleic acid sequences of these sequences are used to inhibit expression  
 CC of a gene from Helicobacter species. H. pylori whole genomic DNA was  
 CC isolated and nebulised to a median size of 2000 bp. Purified DNA  
 CC fragments were blunt-ended and ligated to unique BstXI-linker adapters in  
 CC 100-1000 fold molar excess. These linkers are complementary to the BstXI-  
 CC cut pMPX vectors, while the overhang is not self-complementary. Therefore

CC the linkers will not concatemerise nor will the cut vector re-ligate  
 CC itself easily. The linker-adaptor inserts were ligated to each of the 20  
 CC pMPX vectors to construct a series of shotgun subclone libraries. The  
 CC purified DNA samples were then sequenced. Note: The ORF/protein reference  
 CC number for this sequence was obtained from the related specification,  
 CC W09640893

XX SQ Sequence 253 AA;

## Alignment Scores:

Pred. No.: 6.94e-134 Length: 253  
 Score: 1270.00 Matches: 251  
 Percent Similarity: 99.60% Conservative: 1  
 Best Local Similarity: 99.21% Mismatches: 1  
 Query Match: 94.14% Indels: 0  
 DB: 2 Gaps: 0

US-09-732-091-3 (1-759) x AAW24673 (1-253)

QY 1 ATGGCATCAAAATATGATAGAGACTTGGATTTTAAAGCAATTTGAATCTAGTGATTTA 60  
 Db 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20  
 QY 61 TTGGATTGTTTTCAGGTGCTTGTGTTTGGTAAAGACGCGCAAAACACACAAATCAAAA 120  
 Db 21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40  
 QY 121 CTGACCCAGTCCATAGAAATACAAAGCATCGGCGATGATTAACGCTAATACGCAAGA 180  
 Db 41 LeuThrSerSerIleGluTyrLysArgHisGlyAspTyrAlaLysTyrAlaGluArg 60  
 QY 181 ATCGCTGAACAGTTGCAATACTATGGAGCAATAGTTTTCGAGTTCATTAAAGGCGAA 240  
 Db 61 IleAlaGluGlnLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheLysGlyGlu 80  
 QY 241 GCGAGTCTTATACAAAGAGATTTTATCGCATGTGTGCGATATAATTAAGGTCAATTACA 300  
 Db 81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100  
 QY 301 AAGAAACTGAACAGCTTTAATTAAGCAAAACATGCTTCTAAATCTTAGAAGAGT 360  
 Db 101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120  
 QY 361 TTGGAAGAAATGATGATGAAGAGTGAAGAAATGCGGATGATGATGATGATGATGATGAT 420  
 Db 121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140  
 QY 421 ACGGACAAATTTAAACACACAGCCTTAAAGCGCGCGACTTAAACGCTGTTTAAATGCGG 480  
 Db 141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160  
 QY 481 GGTTTTAAATCTTATCAATAGCTGCTCATTTGTTGCGATGCGTCCGCAAAACCATCTCTA 540  
 Db 161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180  
 QY 541 GCGGCTGGTTCATCGCTTGGCGCAATCAGGTGCTTACAGAACTCTGAGCTTTTAAACA 600  
 Db 181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200  
 QY 601 GGTCTCTGCTCGATCATTACAGCGGTATGGACAGCGATGATGATGATGATGATGATGATGAT 660  
 Db 201 GlyProValGlyTyrPheIleThrGlyValThrAlaIleAspIleAlaGlyProAla 220  
 QY 661 TATAGGGAACCATACCGCATCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 720  
 Db 221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240  
 QY 721 GCCAATGGAGATAAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 759  
 Db 241 AlaAsnGluAspLysSerLeuGlnIleGluSerVal 253

RESULT 5

AAW20866

ID AAW20866 standard; protein; 256 AA.  
 XX AC AAW20866;  
 XX DE  
 XX DT 18-JUL-1997 (first entry)  
 XX H. pylori cytoplasmic protein, 12ge20305orf30.  
 XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 XX identification; binding compound; bacterium; life cycle; activator;  
 XX bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 XX diagnosis.  
 XX OS Helicobacter pylori.  
 XX PN W09640893-A1.  
 XX PD 19-DEC-1996.  
 XX PF 06-JUN-1996; 96WO-US009122.  
 XX PR 07-JUN-1995; 95US-00487032.  
 XX PR 01-APR-1996; 96US-00630405.  
 XX PA (ASTR) ASTRA AB.  
 XX Smith D, Berglindh OC, Mellgaerd BL;  
 XX WPI; 1997-052306/05.  
 XX N-PSDB; AAT68119.  
 XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 XX useful for vaccines to treat or prevent H. pylori infection, and to  
 XX detect Helicobacter.  
 XX Claim 61; Page 1269; 1481pp; English.  
 XX This sequence represents a H. pylori cytoplasmic protein. The protein may  
 XX be used in a vaccine to prevent or treat H. pylori infection or to  
 XX identify H. pylori polypeptide binding compounds, useful as potential H.  
 XX pylori life cycle activators or inhibitors. The genomic sequence of H.  
 XX pylori (ATCC 55679) was determined from overlapping contigs generated by  
 XX mechanically shearing the bacterial DNA. The sequences were analysed for  
 XX ORF of at least 180 nucleotides, and the predicted coding regions defined  
 XX by computer evaluation. To identify likely H. pylori antigens for vaccine  
 XX development, the amino acid sequences predicted from various ORF were  
 XX analysed for significant homology to other known or exported membrane  
 XX proteins. Having identified and determined the sequences of interest,  
 XX particular regions can be isolated from H. pylori by PCR amplification  
 XX for recombinant polypeptide production, e.g. in E. coli hosts  
 XX  
 SQ Sequence 256 AA;  
 Alignment Scores:  
 Pred. No.: 6,99e-134 Length: 256  
 Score: 1270.00 Matches: 251  
 Percent Similarity: 99.60% Conservative: 1  
 Best Local Similarity: 99.21% Mismatches: 0  
 Query Match: 94.14% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-732-091-3 (1-759) x AAW20866 (1-256)  
 QY 1 ATGGCATAAAATATGATGAGACTTGGAAATTTTAAAGCAATTTGGAATCTAGTATT 60  
 Db 4 MetAlaTyrIstYrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 23  
 QY 61 TTGGATTGTTGAGTGCTGTTGTTTGGTAAAGACGGCGAAAGACACATGAA 120  
 Db 24 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 43  
 QY 121 CTGACAGCTCCGATACAAAGCGATGGCGATGATTAACCTAAATACGAGAAAGA 180

Db 44 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 63  
 QY 181 ATGCTGAAGAGTTCCAAATACATTCGGAGCAATAGTTTTCGAGTTTCATTAAGCGGA 240  
 Db 64 IleAlaGluGluLeuGlnTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 83  
 QY 241 GGAGTCTTATACAAAGAGATTTATGCGATGTGCGATAAAATTAAGGTCAATCAAC 300  
 Db 84 GlyValLeuTyrLysGluLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 303  
 QY 301 AAGAAACTGAACCGACTTTAATTCACAAACATCTCTTTTAAATCTTACAGAAAGT 360  
 Db 104 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 123  
 QY 361 TTGGAAGAAATGCGATGATGAAGAGTGAAGAAATGTCGATGAATATCCATAAAAC 420  
 Db 124 LeuGluGluMetAspAspGluValLysGluMetCysAspGluLeuSerIleLysAsn 143  
 QY 421 ACGGACAAATTTAAACAGACAGCCCTTAAGCGCGGACCTTTAAACGCTGTTTAAATGGGG 480  
 Db 144 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaIleThrLeuThrLeuPheLysMetGly 163  
 QY 481 GGTTTTAAATCTTATCAATAGCTGTCTATTGTCGAATCCGTCGCAAAACCATCTTA 540  
 Db 164 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 183  
 QY 541 GGGCGTGTTTATCGTTCCGGGCAATCAGGTGCTTACAGAACTCGAGCTGCTTTTAAACA 600  
 Db 184 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 203  
 QY 601 GGTCTGTGTGGTGGATCATTACAGCGCTATGACAGCGATTCGATTCGAGCGCGGCT 660  
 Db 204 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 223  
 QY 661 TATAGGTAACCATACCGGATGCAATGTGTGTCCTTACGCTTACGCTTAAACACAGCAA 720  
 Db 224 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 243  
 QY 721 GCCAATGGAGATGAAGTGGTTCGAAATAGATCCATT 759  
 Db 244 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 256  
 RESULT 6  
 AAE26880  
 ID AAE26880 standard; protein; 253 AA.  
 XX  
 AC AAE26880;  
 XX  
 XX 13-DEC-2002 (first entry)  
 XX  
 XX Helicobacter sp. pOE/HP30 plasmid DNA encoded protein.  
 XX  
 XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
 XX type B gastritis; anti-inflammatory; adenocarcinoma; defense mechanism;  
 XX low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
 XX immunostimulant; cytostatic; peptic ulcer.  
 XX Helicobacter sp.  
 XX W0200251237-A2.  
 XX  
 XX 04-JUL-2002.  
 XX  
 XX 07-DEC-2001; 2001WO-US048392.  
 XX  
 XX 07-DEC-2000; 2000US-00732091.  
 XX  
 XX (ANTE-) ANTEK BIOLOGICS INC.  
 XX  
 XX Tian J, Walker R, Jackson WJ;  
 XX WPI; 2002-666854/71.  
 XX N-PSDB; AAD44537.  
 XX

XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
PT the proteins, useful as vaccines for raising immune response in animals.  
XX  
XX  
XX Disclosure; Page 126-127; 127pp; English.  
XX  
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and  
CC peptides derived from them. The invention is useful for producing an  
CC immune response. It is useful for preventing, treating or ameliorating a  
CC disorder or disease associated with infection of an animal with  
CC Helicobacter. Pharmaceutical composition and vaccines comprising the  
CC sequences of the invention is useful for treating type B gastritis,  
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B  
CC cell lymphoma. The invention is useful as reagents for clinical or  
CC medical diagnosis of Helicobacter infections and for scientific research  
CC on the properties of pathogenicity, virulence and infectivity of  
CC Helicobacter, as well as host defense mechanisms. The present sequence is  
CC Helicobacter sp. pQE/HP30 plasmid DNA encoded protein  
XX  
SQ Sequence 253 AA;  
  
Alignment Scores:  
Pred. No.: 3.36e-133 Length: 253  
Score: 1264.00 Matches: 250  
Percent Similarity: 99.60% Conservative: 2  
Best Local Similarity: 98.81% Mismatches: 1  
Query Match: 93.70% Indels: 0  
DB: 5 Gaps: 0  
  
US-09-732-091-3 (1-759) x AAE26880 (1-253)  
QY 1 ATGGCATACAAATGATGAGACTTGGAAATTTTAAAGCAATTGGAATCTAGTATTGA 60  
DB 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20  
QY 61 TTGGATTGTTGAGTGCTGTTTGGTAAAGCGCGGAAAGACCAATGAATA 120  
DB 21 LeuAspLeuPheGluValLeuValPheGlyLeuAspGlyGluLysArgHisGlnLys 40  
QY 121 CTGACGAGCTCCATAGATACAAAGCATGGCGATGATGCTAAATACGAGAAAGA 180  
DB 41 LeuThrSerSerLeuGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60  
QY 181 ATCGCTGAAGAGTTGCAATACTATGGAGCAATAGTTTTCGAGTTTTCATTAAGGCGAA 240  
DB 61 IleAlaGluLeuGluGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80  
QY 241 GGAGTCTTATACAAAGAGATTTATGCGATGTCGCGATTAATTAAGCTCAATTAAC 300  
DB 81 GlyValLeuTyrLysGluIleLeuLysCysAspValCysAspLysLeuLysValAsnTyrAsn 100  
QY 301 AAGAAACTCAAGAGCTTTAATTAATGAACAAACATGCTTTCTAAATCTTAGAAGAGT 360  
DB 101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120  
QY 361 TTGAAGAAATGGATGATGAAGAAGTGAAGAAATGCGGATGATTCATTAATTAAGAAAC 420  
DB 121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140  
QY 421 ACGGACAAATTAACACAGACGCTTAAGCGCGCGCTTTAAACGCTGTTTAAATGGG 480  
DB 141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160  
QY 481 GGTTTTAAATCTTATCAATAGCTGTCATTTGCGAATGCGTGCAGAAACCAATCTA 540  
DB 161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180  
QY 541 GGGCGTGTTTATCGCTTGGCGGCAATCAGTGTCTTACAAGAACTCTGAGCTTTTAAACA 600  
DB 181 GlyArgGlyLeuSerLeuAlaGlyAspGlnValLeuThrArgThrLeuSerPheLeuThr 200  
QY 601 GGTCTGTGCTGGATCATTAACAGCGCTATGACAGCGATTTGATTTGACGGGCGGCT 660

DB 201 GlyProValGlyTyrIleIleThrGlyValTyrThrAlaIleAspIleAlaGlyProAla 220  
QY 661 TATAGGGTAACCATACCGCATCATTTGTGGTTCACATTTAGCGCTAAACACACAGCAA 720  
DB 221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240  
QY 721 GCCAATGGAGATAGAGTGTGTCGAAATAGATCCATT 759  
DB 241 AlaAsnGluAspLysSerLeuGlnIleGluSerIle 253  
RESULT 7  
AAE26879  
ID AAE26879 standard; protein; 265 AA.  
AC AAE26879;  
XX  
XX 13-DEC-2002 (first entry)  
DT  
XX Helicobacter sp. pQE/HP56 plasmid DNA encoded protein.  
XX  
XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;  
XX type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;  
XX low grade B cell lymphoma; virulence; antibacterial; gastric cancer;  
XX immunostimulant; cytostatic; peptic ulcer.  
XX  
XX Helicobacter sp.  
XX WO200251237-A2.  
XX  
XX 04-JUL-2002.  
XX  
XX 07-DEC-2001; 2001WO-US048392.  
XX  
XX 07-DEC-2000; 2000US-00732091.  
XX  
XX (ANTE-) ANTEX BIOLOGICS INC.  
XX  
XX Tian J, Walker R, Jackson WJ;  
XX  
XX WPI; 2002-666854/71.  
XX N-PSDB; AAD44536.  
XX  
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding  
PT the proteins, useful as vaccines for raising immune response in animals.  
PS  
PS Disclosure; Page 124-125; 127pp; English.  
XX  
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and  
CC peptides derived from them. The invention is useful for producing an  
CC immune response. It is useful for preventing, treating or ameliorating a  
CC disorder or disease associated with infection of an animal with  
CC Helicobacter. Pharmaceutical composition and vaccines comprising the  
CC sequences of the invention is useful for treating type B gastritis,  
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B  
CC cell lymphoma. The invention is useful as reagents for clinical or  
CC medical diagnosis of Helicobacter infections and for scientific research  
CC on the properties of pathogenicity, virulence and infectivity of  
CC Helicobacter, as well as host defense mechanisms. The present sequence is  
CC Helicobacter sp. pQE/HP56 plasmid DNA encoded protein  
XX  
SQ Sequence 265 AA;  
  
Alignment Scores:  
Pred. No.: 3.36e-133 Length: 265  
Score: 1264.00 Matches: 250  
Percent Similarity: 99.60% Conservative: 2  
Best Local Similarity: 98.81% Mismatches: 1  
Query Match: 93.70% Indels: 0  
DB: 5 Gaps: 0  
  
US-09-732-091-3 (1-759) x AAE26879 (1-265)  
QY 1 ATGGCATACAAATATGATGAGACTTGGAAATTTTAAAGCAATTGGAATCTAGTATTGA 60



```

AC AAE26876;
XX
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #5.
XX
XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
XX WO200251237-A2.
XX
XX 04-JUL-2002.
XX
XX 07-DEC-2001; 2001WO-US048392.
XX
XX 07-DEC-2000; 2000US-00732091.
XX
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Tian J, Walker R, Jackson WJ;
XX
XX WPI; 2002-666854/71.
XX
XX N-PSDB; AAD44529.
XX
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
XX Claim 7; Page 116; 127pp; English.
XX
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
XX Sequence 49 AA;
SQ
Alignment Scores:
Pred. NO.: 2.12e-19 Length: 49
Score: 252.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.68% Indels: 0
DB: 5 Gaps: 0
US-09-732-091-3 (1-759) x AAE26876 (1-49)
QY 607 GTTGCTGGATCATTACAGCGGTATGGACAGGATTGATATTGACGGCGGCTTATAGG 666
Db 1 ValGlyThrIleThrGlyValTrpThrIleAlaAspIleAlaGlyProAlaTyraG 20
QY 667 GTAAACCATACCGGCGATGTTGGTTGCCACTTTACGCTTAAACACAGCAGCCAAAT 726
Db 21 ValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnAlaAsn 40
QY 727 GGAGATAGAGAGTGGTTCGAATAGAA 753
Db 41 GlyAspLysLysSerLeuGlnIleGlu 49
RESULT: 10
AAE26873
ID AAE26873 standard; peptide; 38 AA.

```

```

XX AAE26873;
AC
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #2.
XX
XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
XX WO200251237-A2.
XX
XX 04-JUL-2002.
XX
XX 07-DEC-2001; 2001WO-US048392.
XX
XX 07-DEC-2000; 2000US-00732091.
XX
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Tian J, Walker R, Jackson WJ;
XX
XX WPI; 2002-666854/71.
XX
XX N-PSDB; AAD44526.
XX
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
XX Claim 7; Page 115; 127pp; English.
XX
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
XX Sequence 38 AA;
SQ
Alignment Scores:
Pred. NO.: 2.34e-13 Length: 38
Score: 198.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.68% Indels: 0
DB: 5 Gaps: 0
US-09-732-091-3 (1-759) x AAE26873 (1-38)
QY 157 GATTACGCTAAATACGCAGAAAGAAATCGCTGAAGAGTTGCAATACTATGGAGCAATAGT 216
Db 1 AspTyraIleLysTyraIleAlaGluArgIleAlaGluLeuGlnTyTyTyGlySerAsnSer 20
QY 217 TTTCGAGCTTTTCATTAAAGGCGAAGGAGCTCTTATACAAAGAGATTTTATCGAT 270
Db 21 PheAlaSerPheIleLysGlyGluGlyValLeuTyTyLysGluIleLeuCysAsp 38
RESULT: 11
AAE26875
ID AAE26875 standard; peptide; 41 AA.
XX
XX AAE26875;
XX

```

```

DT 13-DEC-2002 (first entry)
XX
XX Helicobacter sp. HP30-derived peptide #4.
XX
XX HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
XX type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
XX low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
XX immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
XX WO200251237-A2.
XX
XX 04-JUL-2002.
XX
XX 07-DEC-2001; 2001WO-US048392.
XX
XX 07-DEC-2000; 2000US-00732091.
XX
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Tian J, Walker R, Jackson WJ;
XX WPI; 2002-666854/71.
XX N-PSDB; AAD44528.
XX
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
XX the proteins, useful as vaccines for raising immune response in animals.
XX
XX Claim 7; Page 116; 127pp; English.
XX
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and
XX peptides derived from them. The invention is useful for producing an
XX immune response. It is useful for preventing, treating or ameliorating a
XX disorder or disease associated with infection of an animal with
XX Helicobacter. Pharmaceutical composition and vaccines comprising the
XX sequences of the invention is useful for treating type B gastritis,
XX peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
XX cell lymphoma. The invention is useful as reagents for clinical or
XX medical diagnosis of Helicobacter infections and for scientific research
XX on the properties of pathogenicity, virulence and infectivity of
XX Helicobacter, as well as host defense mechanisms. The present sequence is
XX Helicobacter sp. HP30-derived peptide
XX
XX Sequence 41 AA;

Alignment Scores:
Pred. No.: 6.82e-13 Length: 41
Score: 194.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.38% Indels: 0
DB: 5 Gaps: 0

US-09-732-091-3 (1-759) x AAE26875 (1-41)
QY 433 AACAGCAGCCTTAGCGGGGACTTTAAGCTCTTTAAATGGGGGTTTAAATCT 492
DB 1 AsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGlyGlyPheLysSer 20
QY 493 TATCAATTAGCTGTCATTGTGGAAATGCGGTGCGCAAAACCACTTCTAGGGCGTGGTTTA 552
DB 21 TyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleuGlyArgGlyLeu 40
QY 553 TCG 555
DB 41 Ser 41
RESULT 12
AAE26874
ID AAE26874 standard; peptide; 30 AA.
XX
XX AAE26874;
AC

```

```

DE Helicobacter sp. HP30-derived peptide #1.
XX
XX HP30; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
XX WO200251237-A2.
XX
XX 04-JUL-2002.
XX
XX 07-DEC-2001; 2001WO-US049392.
XX
XX 07-DEC-2000; 2000US-00732091.
XX
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Tian J, Walker R, Jackson WJ;
XX
XX WPI; 2002-666854/71.
XX
XX N-PSDB; AAD44525.
XX
XX Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
PT
XX
XX Claim 7; Page 1:5; 127pp; English.
XX
XX The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
XX Sequence 30 AA;
SQ
Alignment Scores:
Pred. No.: 5.5e-08 Length: 30
Score: 150.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.12% Indels: 0
DB: Gaps: 0

US-09-732-091-3 (1-759) x AAE26872 (1-30)
QY 1 ATGGCATACAAATATCATAGAGACTTGGAAATTTTAAAGCAATTTGGATCTAGTGATTTA 60
Db 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20
QY 61 TTGGATTTGTTGAGGTGCTGTTTTTGGT 90
Db 21 LeuAspLeuPheGluValLeuValPheGly 30

RESULT 14
AAW44365
ID AAW44365 standard; peptide; 173 AA.
XX
XX AAW44365;
AC
XX 28-MAY-1998 (first entry)
DT
XX Breast cancer-associated protein fragment BC-2 SEQ ID NO:12.
DE
XX Human; breast cancer-associated protein; nuclear matrix protein;
KW

```

detection; diagnosis; antibody.

Homo sapiens.

WO9746884-A1.

11-DEC-1997.

03-JUN-1997; 97WO-US009529.

05-JUN-1996; 96US-00658639.

(MATR-) MATRITECH INC.

Keese SK, Obar R, Wu Y;

WPI; 1998-042336/04.

N-PSDB; AAV15304.

Diagnosing breast cancer by detecting a breast cancer-associated protein  
PT - allows early and reliable diagnosis and treatment monitoring, and  
PT antibody or inhibitory compounds useful for treating breast cancer.

Claim 11; Page 32; 47pp; English.

The present sequence represents a breast cancer-associated protein  
CC fragment for use in a method for diagnosing breast cancer in tissue or  
CC body fluid by detecting one or more breast cancer-associated protein(s).  
CC Alternatively a nucleic acid encoding a breast cancer-associated protein  
CC is detected in the sample by the use of a nucleic acid probe. The breast  
CC cancer-associated proteins, encoding nucleic acids and antibodies are  
CC useful in diagnostic assays and kits for breast cancer detection. The  
CC proteins are also useful in screening for inhibitory compounds and  
CC monitoring effectiveness of treatments. The antibody, or pharmaceutical  
CC compositions containing the antibody or the inhibitory compound, can be  
CC administered to patients to treat breast cancer. The methods allow  
CC reliable and early diagnosis of breast cancer, or prediction of its  
CC onset, by detection of specific markers expressed in breast tumour cells  
CC but not in non-cancerous breast cells

Sequence 173 AA;

Alignment Scores:

Pred. No.: 0.0325 Length: 173  
Score: 101.50 Matches: 41  
Percent Similarity: 43.58% Conservative: 37  
Best Local Similarity: 22.91% Mismatches: 72  
Query Match: 7.52% Indels: 29  
DB: Gaps: 5

US-09-732-091-3 (1-759) x AAW44365 (1-173)

QY 26 TGGAAATTTTAAAGCAATTCGAATCTAGTGATTTATTTGGATTTGTTGAGTGTGTTT 85

Db 2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp----- 14

QY 86 TTGGTAAAGACGGCGGAAAGAACACATCGAATAAAGCTGACCGCTCCATAGATACAAAA 145

Db 15 -----GlnLysThrTrpValAlaProGlyTyrValArg 25

QY 146 GGCATGGCGATGATTACGCTAAATAACGAGAAAGATCGCTGAAGAGTTTGCATACTATG 205

Db 26 LysPheValLeuMetArgAlaAsnIleGlnAlaValSerLeuLys-IleGlnThrLeuL 45

QY 206 GGAGC---AATAGTTTGGAGTTTTCATTTAAAGGC-----GAAGGAGTCTTAT 250

Db 45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65

QY 251 ACAAGAGATTTTATCGCATGTGTCGATATAAATTAAGGTCAATTACACAGAAACTG 310

Db 65 snArgGlnLeuLysLeuProGlnIleGlnLysIleMetMetGluPheGluArgGlnAlaG 85

QY 311 AAACGACTTTAATTGAACAAACATGCTTTCTAAATCTTTAGAAAGAGTTTGGAGAAA 370

```
Db      85 luilleMetAspMetLysGluMetMetAsnAspAlaileAspAspAlaMetGlyAspG 105
QY      371 TGGATGATGAAGAA-----GTGAAGAANTGTGCGATGAATATTCNTAA 415
Db      105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeuS 125
QY      416 AAAACACGACCAATTTAAACGACCAAGCTTAAGCGCGCGACCTTTAAAGCGCTGTTTAA 475
Db      125 erLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeuSerVal---Ala 144
QY      476 TGGGGGGTTTAAATCTATCAATAGCTGCTCATTTGCGAATGCGTGCSCA 528
Db      144 laGlyGlyLysAlaGluAlaAlaSerAlaLeuAlaAspAlaAspAla 161

RESULT 15
AAB98717
ID AAB98717 standard; protein; 173 AA.
AC AAB98717;
XX
XX 29-AUG-2001 (first entry)
DT Human breast cancer-associated protein BC-2 portion, SEQ ID NO:12.
DE Human breast cancer-associated protein; BC-2; marker protein;
XX nuclear matrix protein; diagnosis; detection; tumour.
XX Homo sapiens.
XX USG218131-B1.
XX 17-APR-2001.
XX
XX 06-OCT-1997; 97US-00944604.
XX
XX 05-JUN-1996; 96US-00658639.
XX (MATR-) MATRITECH INC.
XX
XX Keesee SK, Obar R, Wu Y;
XX WPI; 2001-396355/42.
XX N-PSDB; AAB25091.
XX
```

Diagnosing or detecting breast cancer in an individual comprises detecting the presence of breast cancer-associated proteins in a biological sample.

Example 1; Col 21-24; 30pp; English.

The invention relates to novel human breast cancer-associated proteins, and their use in diagnosing and detecting breast cancer. The breast cancer-associated proteins of the invention are nuclear matrix proteins designated BC-2 (AAB98720), BC-8 isoform A (BC-8A, AAB98721) and CC BC-8 isoform B (BC-8B, AAB98722), although 6 other breast cancer-associated nuclear matrix proteins (BC-1, BC-3, BC-4, BC-5, BC-6, and BC-7) were also isolated (sequences not given in the specification). The novel breast cancer-associated proteins are present in the nuclear matrix of breast cancer cells, but are not present in the nuclear matrix of cells from normal breast tissue. The invention also encompasses fragments of the breast cancer-associated proteins (AAB98709-AAB98716), and methods for their use in breast cancer diagnosis. The breast cancer marker proteins of the invention or fragments thereof may be detected in a sample of breast tissue or blood from an individual. Isolated breast cancer-associated proteins may also be used to prepare antibodies, particularly monoclonal antibodies, which may be used to detect or diagnose breast cancer. Nucleic acids encoding the breast cancer-associated proteins may also be used in the diagnosis and detection of breast cancer, and in the isolation of DNA or protein sequences which may interact with breast cancer-associated nuclear matrix proteins. The present sequence represents a portion of breast cancer-associated protein BC-2 referred to in an exemplification of the invention

```
XX SQ Sequence 173 AA;
Alignment Scores:
Pred. No.: 0.0325 Length: 173
Score: 101.50 Matches: 41
Percent Similarity: 43.58% Conservative: 37
Best local Similarity: 22.91% Mismatches: 72
Query Match: 7.52% Indels: 29
DB: 4 Gaps: 5

US-09-732-091-3 (1-759) x AAB98717 (1-173)
QY 25 TGGAAATTTTAAAGCAATGGAATCTAGTCAATTTATGCAATGTTTGGAGTCTCTGTTT 85
Db 2 TtpProSerLysAlaArgTfpMetLeuPheAlaSerTip----- 14
QY 86 TTGGTAAGACGCGGGAAGAAACACACATGAAAACTGACCGCTCCATAGATACAAA 145
Db 15 -----GlnLysThrTfpValAlaProGlyTyrValArg 25
QY 146 GGCATGGCGATGATTACGCTAAATACGCAGAAAGATCGCTGAAGAGTTGCAATACTATG 205
Db 26 LysPheValLeuMetArgAlaAsnIleGlnAlaValSerLeuLys---IleGlnThrLeuL 45
QY 206 GGAGC---AATAGTTTTCGAGTTTCATTAAGGC-----GAAGGAGTCTTAT 250
Db 45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65
QY 251 ACAAGAGATTTTATCGCATGTGTGCGATAAATTAAGGTCAATTACACAAAGAACTG 310
Db 65 snArgGlnLeuLysLeuProGlnIleGlnLysIleMetMetGluPheGluArgGlnAlaG 85
QY 311 AAACGACTTTAATTGAACAAAACATGCTTCTTAAATCTTAGAAAGAGTTTGGAGAGAA 370
Db 85 luileMetAspMetLysGluGluMetMetAsnAspAlaileAspAspAlaMetGlyAspG 105
QY 371 TGGATGATGAAGAA-----GTGAAGAANTGTGCGATGAATATTCNTAA 415
Db 105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuS 125
QY 416 AAAACACGACCAATTTAAACGACCAAGCTTAAGCGCGCGACCTTTAAAGCGCTGTTTAA 475
Db 125 erLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeuSerVal---Ala 144
QY 476 TGGGGGGTTTAAATCTATCAATAGCTGCTCATTTGCGAATGCGTGCSCA 528
Db 144 laGlyGlyLysAlaGluAlaAlaSerAlaLeuAlaAspAlaAspAla 161
```

Search completed: July 5, 2004, 03:03:13  
Job time : 69 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 5, 2004, 03:00:58 ; Search time 22.5 Seconds

(without alignments)  
6489.722 Million cell updates/sec

Title: US-09-732-091-3

Perfect score: 1349  
Sequence: 1 atggcatacaaatatgatag.....cggtgcaaatagaatccatt 759

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09732091/runat\_02072004.182407.19036/app query.fasta\_1.903  
-DB=PIR\_78 -QMT=fastan -SURF=n2p.rpr -MINMATCH=0.1 -LOOEXT=0 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR\_SCORES=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US09732091.gcn 1.1.25 @runat\_02072004.182407.19036 -NCPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR 78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1279	94.8	253	D64718	conserved hypothe
2	1270	94.1	253	B71800	hypothetical prote
3	784.5	58.2	207	E64718	conserved hypothe
4	630	46.7	209	A71800	hypothetical prote
5	472.5	35.0	155	C64718	conserved hypothe
6	265	19.6	237	C64721	hypothetical prote
7	265	19.6	237	AC0503	conserved hypothe
8	262	19.4	237	C90630	probable oxidoredu
9	262	19.4	237	C85481	probable oxidoredu
10	133	9.9	39	F64718	hypothetical prote
C 11	127.5	9.5	206	D90630	hypothetical prote
C 12	127.5	9.5	206	D85481	hypothetical prote
C 13	122	9.1	196	A40623	heat shock protein
14	98.5	7.3	1009	C89910	hypothetical prote

15	98	7.3	282	2	A80360	hypothetical prote
16	98	7.3	1099	2	G90546	conserved hypothe
17	97.5	7.2	311	2	C81380	probable D-2-hydro
18	97.5	7.2	728	2	A81385	probable ATP /GTP
C 19	97	7.2	346	2	A70144	hypothetical prote
20	97	7.2	916	2	E71330	probable preprotei
21	96.5	7.2	373	2	G70355	8-amino-7-oxonan
22	96	7.1	262	2	T33071	hypothetical prote
23	96	7.1	420	2	F96920	probable permease
24	95.5	7.1	344	2	A47025	abortive phage res
25	95	7.0	517	2	C81327	hypothetical prote
26	95	7.0	2649	2	A40937	bullous pemphigoid
27	94.5	7.0	387	2	A86302	hypothetical prote
28	93.5	6.9	802	2	S45252	penicillin amidase
29	93	6.9	355	2	E90488	transposase ISC122
30	93	6.9	610	2	G70178	exodeoxyribonucle
31	92.5	6.9	430	2	T33070	hypothetical prote
32	92.5	6.9	802	2	I39665	penicillin amidase
33	92.5	6.9	886	2	T16536	hypothetical prote
34	92	6.8	308	2	H81345	hypothetical prote
35	92	6.8	1939	2	T18372	repeat organellar
36	91.5	6.8	399	2	G72253	RNA polymerase sag
37	91.5	6.8	472	2	A64320	PER112 homolog - M
38	91.5	6.8	508	2	T24622	hypothetical prote
39	91.5	6.8	515	2	S42797	rad 26 protein - f
40	91.5	6.8	1769	2	S53378	probable membrane
41	91.5	6.8	1847	2	E64477	replication factor
42	91	6.7	284	2	A45488	body-wall muscle t
43	91	6.7	359	2	A12368	hypothetical prote
44	90.5	6.7	320	2	B97214	hypothetical prote
45	90.5	6.7	440	2	E64571	transposase-like p

#### ALIGNMENTS

RESULT  
D64718

conserved hypothetical protein [HP1588] - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C:Accession: D64718

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glöck, A.; McKen  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,  
Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: D64718

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-253 <TOM>

A:Cross-references: GB:AE000656; GB:AE000511; NID:G2314771; PIDN:AAD08627.1; PID:G2314

#### Alignment Scores:

Pred. No.:	1-5e-97	Length:	253
Score:	1279.00	Matches:	253
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	94.81%	Indels:	0
DB:	2	Gaps:	0

US-09-732-091-3 (1-759) x D64718 (1-253)

QY	1	ATGGCATCAATATGATAGACTTGAATTTTAAAGCAATCGAATCTAGTATTTA 60
DB	1	MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20
QY	61	TTCGATTTTTCAGCTGCTCTTTTTCGTAAGCGCGGAAAGACACATGAAAAA 120
DB	21	LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGlyLysArgHisnGluLys 40
QY	121	CTGACCAGCTCCATAGAAATACAAAGGCGATGCGGATGATTACGCTAAATACGCAAGAAGA 180



Alignment Scores:  
 Pred. No.: 8,01e-57 Length: 207  
 Score: 784.50 Matches: 153  
 Percent Similarity: 83.65% Conservative: 21  
 Best Local Similarity: 73.56% Mismatches: 33  
 Query Match: 58.15% Indels: 1  
 DB: 2 Gaps: 1

US-09-732-091-3 (1-759) x E64718 (1-207)

QY 121 CTGACCTCCATAGATCAAAAGCGCATGCGGATGATCGCTAAATACCGAAGA 180  
 Db 1 MetThrSerSerThrGluThrGlnArgTyrGlyTyrAspCysAlaLysTyrProArg 20  
 QY 181 ATCGCTGAAGAGTTCGATATCATCGGAGCAATAGTTTTCGAGTTCATTAAGCGCAA 240  
 Db 21 IleAlaGluLeuGlnArgTyrGlyGlyAsnSerPheMetAsnPheArgAspGlu 40  
 QY 241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGCGATTAATTAAGTCAATTACAC 300  
 Db 41 GlyValLeuTyrLysGluLeuLeuCysAspAlaCysAspHisLeuLysValAsnTyrAsn 60  
 QY 301 AAGAAACCTGAACAGACTTAATTAAGCAACAAACATGCTTCTAAATCTTAGAAGAGT 360  
 Db 61 LysLysSerAspThrThrLeuLeuGluAsnMetLeuSerSerIleLeuGlnLysSer 80  
 QY 361 TTGGAAGAAATGATGATGAAGAGTGAAGAAATGTGCGATGAATTTATCCATAAAAC 420  
 Db 81 LeuGluLysMetSerAspGluGluLeuArgGluLeuCysAspGluLeuGlyValLysAsn 100  
 QY 421 ACGGACATTTAAACAGACAGACCTTAAGCGCGGACCTTTAACGCTGTTTAAATGGG 480  
 Db 101 ThrAsnLysLeuGlyLysGlnAlaLeuSerThrAlaAlaLeuThrLeuPheArgMetGly 120  
 QY 481 GGTTTTAATCTTATCAATTAAGTCTGATGTTGCGATGCGTGCAGAAACCACTTCA 540  
 Db 121 GlyPheLysSerTyrGlnLeuAlaLeuIleValAlaAsnAlaValIleLysAlaLysPhe 140  
 QY 541 GGGCGTGTGTTATCCCTTCGCGGCAATCAGGTGCTTACAAAGAACTCTGAGCTTTTAA 600  
 Db 141 GlnArgGlyLeuSerLeuGlyAlaAsnAlaAlaLeuThrArgGlyLeuSerIleLeuThr 160  
 QY 601 GGTCTGTTGCTGATCATTAACAGCGTATGCGATGCGATGATGATGCGCGCGGT 660  
 Db 161 GlyProIleGlyTyrIleThrGlyValTyrThrAlaIleAspIleAlaGlyProAla 180  
 QY 661 TATAGGTAACCATACCGCATGATGTTGTTGCGCTTACGCTTAAACACACAGCAA 720  
 Db 181 TyrArgValThrIleProAlaCysIleLeuValAlaThrLeuArgLeuLysAla---Gln 199  
 QY 721 GCCAATGAGATGAAGATGCGGTG 744  
 Db 200 AlaAsnGluIleLysAsnIleLeu 207

RESULT 4  
 A71800  
 hypothetical protein jhp1493 - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
 C:Accession: A71800  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
 Ives, C.; Gibson, R.; Merberg, D.; Malls, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
 A:Reference number: A71800; NID:99120557; PMID:99233682  
 A:Accession: A71800  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-209 <ARN>  
 A:Cross-references: GB:AE001571; GB:AE001439; NID:G4156120; PIDN:AAD07072.1; PID:G415612  
 A:Experimental source: strain J99

C:Genetics:  
 A:Gene: jhp1493

Alignment Scores:  
 Pred. No.: 4.25e-44 Length: 209  
 Score: 630.00 Matches: 122  
 Percent Similarity: 75.49% Conservative: 32  
 Best Local Similarity: 59.80% Mismatches: 46  
 Query Match: 46.70% Indels: 4  
 DB: 2 Gaps: 2

US-09-732-091-3 (1-759) x A71800 (1-209)

QY 112 AATGAAATCTGACGCGCTCCATAGATCAAAAGCGCATGCGGATGATTCAGCTAATAC 171  
 Db 2 AsnGluGluLeuThrSerLeuThrGluTyrGlnArgTyrGlyHisAspTyrAlaLysTyr 21  
 QY 172 GCAGAAAGAAATCGCTGAAGAGTTCGAATACTATGGAGCAATAGTTTTCGAGTTCATT 231  
 Db 22 ProArgArgIleAlaGluLeuGlnArgTyrGlyGlyAsnSerPheAlaAsnPhePhe 41  
 QY 232 AAAGCGGAGAGTCTTATCAAAAGAGATTTTATGCGATGCTGCGATTAATTAAGTCA 291  
 Db 42 ArgAspGluGlyValLeuTyrLysGluLeuCysAspAlaCysAspHisLeuAspIle 61  
 QY 292 AATTACACACAGAAACTGAAAGACTTTAATTAAGCAAAACATGCTTCTTAAATCTTA 351  
 Db 62 AsnTyrAsnGluArgSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81  
 QY 352 GAAAGAAATGCGAAGAAATGATGATGAAGAGTGAAGAAATGTGCGATGAATATATCC 411  
 Db 82 LysAspSerLeuGluLysMetSerGlyArgGluIleLysGluLeuCysAspGlyLeuGly 101  
 QY 412 ATATAAAACACGACCAATTTA-----ACAGACAAAGCCTTAAGCGCGGCGACTTTA 462  
 Db 102 MetProAsnIleAspLysValIleGlyGluAsnLysGlnValLeuIleAlaSerValLeu 121  
 QY 463 AGCTGTTTAAATGGGGGTTTAAATCTTATCAATAGCTGCTGCTGCTGCTGCTGCTG 522  
 Db 122 ThrLeuPheLysAlaGlyLysSerHisSerTyrAlaLeuAlaValAlaAlaAlaAla 141  
 QY 523 GTCGCAAAACCATTCCTAGGCGGTGTTA---TCGCTTGGCGGCAATCAGGTGCTTACA 579  
 Db 142 MetValArgGlnThrLeuGlyHisGlyLeuSerSerValValGlyLysValAlaLeuLys 161  
 QY 580 AGAAGCTCTGAGCTTTTAAACAGCTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 639  
 Db 162 LysThrLeuAspIleLeuAlaGlyProIleGlyTyrValIleThrGlyAlaLeuValSer 181  
 QY 640 ATTGATATTGACGGCGCGCTTATAGGTAACCATACCGCATGCTGCTGCTGCTGCTGCT 699  
 Db 182 IleAsnLeuAlaGlyProAlaTyrArgValThrValProAlaCysValLeuValAlaThr 201  
 QY 700 TTACGCGCTAAAA 711  
 Db 202 LeuArgLysLys 205

RESULT 5  
 C64718  
 conserved hypothetical protein HPI587 - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.;  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaklax, H.G.; Glodek, A.; Mcker  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,  
 Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: C64718  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA



A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: A3C502; MUID:21534947; PMID:11677608  
A:Accession: AC0503  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-237 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD01163.1; PID:GL6501293; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY0010

Alignment Scores:  
Pred. No.: 4,92e-14 Length: 237  
Score: 265.00 Matches: 73  
Percent Similarity: 49.1% Conservative: 40  
Best Local Similarity: 32.64% Mismatches: 101  
Query Match: 19.64% Indels: 22  
DB: 2 Gaps: 7

US-09-732-091-3 (1-759) x AC0503 (1-237)

```
QY 1 ATGGCATCAATATGATAGAGCTTGAATTTTAAAGCAATGGAATCTAGTGAATTA 60
D 3 ValThrTyrLeuHisAspGluLeuAspPheLeuGlnHisCysSerGluGlnLeu 22
QY 61 TTGGATTTGTTGAGGCTGCTTTTGGTAAAGACGGCGAAGAAAGA----- 108
D 23 AlaaspPheAlaArgLeuLeuThrHisasnGluLysGlyLysAlaArgLeuSerVal 42
QY 109 -----CACAAATGAAATACACAGAGCTCCATAGAAATACAAAGGCATGGCGATTA 162
D 43 LeuSerHisasnGlu-----LeuPheLysAlaMetGlu-----GlyHisProGluGlnHis 59
QY 163 GCTAAATACGAGAAAGAAATCGCTGAAGAGTTGCAATCTATGAGGCAATAGTTTGG 222
D 60 ArgArgAsnTyrGlnLeuIlealaglyGluPheGlnHisTyrGlyaspSerIleAla 79
QY 223 AGTTTCATTAAGCGAAGAGTCTTATACAAAGAGATTNTGCGATGTGTGCGATAA 282
D 90 AsnLysLeuArgGlyHisGlyLysGlnTyrArgAlaIleLeuLeuAspValAlaLysArg 99
QY 283 TTAAGGTCAATATACAAAGAAACTGAAACGACTTTAATGAAACAAACATGCTTCT 342
D 100 LeuLysLeuLysAlaAspLysSerMetSerThrPheGluIleGlnGlnLeuLeuGlu 119
QY 343 AAAATCTTGAAGAGTTGGAAGAAATGATGATGATGATGATGATGATGATGATGAT 402
D 120 HisPheLeuArgHisThrTyrGlnLysMetAspAlaAlaHisLysGlnGlnPheLeuGln 139
QY 403 GAATTTATCCATAAAGAACACGACAAATTTAAACGACAGCCTTAAGCGCGCGACTT 462
D 140 AlaValAspAlaLysValSerGluLeuGluLeuLeuProLeuLeuMetLysAspArg 159
QY 463 ACGCTGTTTAA-----ATCGGGGGTTTAAATCTTATCAATAGCTGTGCTTTCGGAAT 519
D 160 SerLeuAlaLysGlyValSerHisLeuLeuSerThrGlnLeuThrArgIleLeuArgThr 179
QY 520 CGGTGCGAAACACCTTCTAGGGGTGCTTATCGCTTTCGCGGCAATCAGGTGCTTACA 579
D 180 HisAlaAlaMetSerIleLeuGlyHisGlyLeu-----LeuArgGly----- 193
QY 580 AGAACTCTGAGCTTTTAAACAGGCTCTGTGTGCTGATCATTACAGCGGTATGACAGCG 639
D 194 -----AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsnGlyValLysAla---- 209
QY 640 ATTGATATTGCGGCGCGCTTATAGGGTAACCATACCGGCAATGATGTTGGTTCCTACT 699
D 210 -----XetSerGlySerAlaTyrArgValThrIleProAlaValLeuGlnIleAlaCys 227
QY 700 TTACGC 705
D 228 LeuArg 229
```

RESULT 8

C90630  
probable oxidoreductase EC3012 [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C90630  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90630  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-237 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA033434.1; PID:G13359467; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: EC50012

Alignment Scores:  
Pred. No.: 8,69e-14 Length: 237  
Score: 262.00 Matches: 76  
Percent Similarity: 48.9% Conservative: 45  
Best Local Similarity: 30.7% Mismatches: 90  
Query Match: 19.42% Indels: 36  
DB: 2 Gaps: 6

US-09-732-091-3 (1-759) x C90630 (1-237)

```
QY 7 TACAATATGATAGAGCTTGAATTTTAAAGCAATGGAATCTAGTGAATTTATGGAT 66
D 5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGlnLeuAlaAsn 24
QY 67 TTGTTGAGGTCCTGTTTTCGTAAAGACGGCGAAGAAAGACAACTAAAAAATCGACC 126
D 25 PheAlaArgLeuLeuThrHisasnGluLysGlyLysThrArgLeuSerSerValLeuMet 44
QY 127 ACCTCCATAGATACAAAGG-----CATGGCGATGATTACGCTAATAACGACGAA 177
D 45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgAsnTyrGln 64
QY 178 AGAATCGCTGAAGAGTTGCAATCTATGAGGCAATAGTTTTCGAGTTCATTAAAGGC 237
D 65 LeuIleAlaGlyGluLeuGlnHisPheGlyLysPheGlyLysPheGlyLysPheGly 84
QY 238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAGGTCAATTAC 297
D 85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104
QY 298 AACAGAAACCTGAACGCTTTAATTTGAACAAACATGCTTTCTTAAATCTTAGAAGA 357
D 105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGlnPheLeuArgAsn 124
QY 358 AGTTTGAAGAAATCGATGATGAA-----GAA 384
D 125 ThrTyrLysLysMetAspGluGlnHisLysGlnGluPheLeuHisAlaValAspAlaArg 144
QY 385 GTGAAGAAATGCGCATGAATTTATCCATAAAGAAACACGACAAATTTAAACGACAGGC 444
D 145 ValasnGluLeuGluLeuLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164
QY 445 TTAAGCGCGGCGACTTTAAGCTGTTTAAATGGGGGTTTAAATCTTATCAATAGCT 504
D 165 ValSer-----HisLeuLeuSerSerGlnLeuThr 174
QY 505 GTCAATGTTGCAATCGCTCGCAAAACCATTTCTAGGCGGTGTTTATCGCTTGGCGG 564
D 175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu-----LeuArgGly 193
QY 565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTAAACAGGTCTCTGTTGGCTGATCATACA 624
D 194 -----AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205
```

QY 625 GCGTATGGACGCGATTGATATTCAGCGCGCGCTTATAGGTTAAACATACCGGCATCC 684  
 |||||  
 Db 206 GlyValLysAla-----ValSerGlySerSerTyrArgValThrIleProAlaVal 222  
 |||||  
 QY 685 ATTGGTGGTGCACCTTTACGC 705  
 |||||  
 Db 223 LeuGlnIleAlaCysLeuArg 229  
 |||||

RESULT 9  
 C85481  
 probable oxidoreductase Z0011 [imported] - Escherichia coli (strain O157:H7, substrain B)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: C85481  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamocousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: C85481  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-237 <STO>  
 A:Cross-references: GB:AE005174; NID:g12512689; PIDN:AAG54311.1; GSPDB:GN00145; UWGP:Z00  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: Z0011

Alignment Scores:  
 Pred. No.: 8,69e-14 Length: 237  
 Score: 262.00 Matches: 76  
 Percent Similarity: 48.99% Conservatives: 45  
 Best Local Similarity: 30.77% Mismatches: 90  
 Query Match: 19.42% Indels: 36  
 DB: 2 Gaps: 6

## US-09-732-091-3 (1-759) x C85481 (1-237)

QY 7 TACAAATATGACGACTTGGAAATTTTAAAGCAATTCAGTATTTATGGAT 66  
 |||||  
 Db 5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGlnLeuAlaAsn 24  
 |||||  
 QY 67 TTGTTGAGTGTCTGTTTGGTAAAGACGCGCGGAAACACACATGAAACCTGACC 126  
 |||||  
 Db 25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerValLeuMet 44  
 |||||  
 QY 127 AGCTCCATAGATACAAAGG-----CATGGCGATGATACGCTAAATACGCAGAA 177  
 |||||  
 Db 45 ArgAspGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgAsnTrpGln 64  
 |||||  
 QY 178 AGAATCGCTGAAGAGTTGCAATACTATGGAGCAATAGTTTTCGAGTTTTCATTAAGGC 237  
 |||||  
 Db 65 LeuIleAlaGlyGluLeuGlnHisPheGlyAspSerIleAlaAsnLysLeuArgGly 84  
 |||||  
 QY 238 GAAGGAGCTTTATACAAAGAGATTTTATCGATGTGCGCATTAATTAAGGTCATATAC 297  
 |||||  
 Db 85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuAla 104  
 |||||  
 QY 298 AACAGAAACTGAAACGACTTTAAATGMAACAAACATGCTTTCTAAATCTTAGAAGA 357  
 |||||  
 Db 105 AspLysGluMetSerThrPheGluIleGluGlnLeuLeuGluGlnPheLeuArgAsn 124  
 |||||  
 QY 358 AGTTTGAAGAATAATGATGATGAA-----GAA 384  
 |||||  
 Db 125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144  
 |||||  
 QY 385 GTGAAGAAGATGTCGATGAATTTATCCATTAACACGACATTTAAACACAGACGCC 444  
 |||||  
 Db 145 ValAsnGluLeuGluLeuLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164  
 |||||  
 QY 445 TTAAGCGCGGACTTTAAACGCTGTTTAAATGGGGGTTTAAATCTTATCAATTAGCT 504  
 |||||  
 Db 165 ValSer-----HisLeuLeuSerSerGlnLeuThr 174

QY 505 GTCATTTGCGAATCGCGTGCAGAAACATTTCTAGGCGGTGGTTTATCGTTGCGGCG 564  
 |||||  
 Db 175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193  
 |||||  
 QY 565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTAAACAGAGTCCTGCTGGCTGGATCATACA 624  
 |||||  
 Db 194 -----AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205  
 |||||  
 QY 625 GCGTATGGACGCGATTGATATTCAGCGCGCGCTTATAGGTTAAACATACCGGCATGC 684  
 |||||  
 Db 206 GlyValLysAla-----ValSerGlySerSerTyrArgValThrIleProAlaVal 222  
 |||||  
 QY 685 ATTGGTGGTGCACCTTTACGC 705  
 |||||  
 Db 223 LeuGlnIleAlaCysLeuArg 229  
 |||||

## RESULT 10

F64718  
 hypothetical protein HP1590 - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 C:Accession: F64718  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McK  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,  
 Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: F64718  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-39 <TOM>  
 A:Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08629.1; PID:g231

Alignment Scores:  
 Pred. No.: 0.00363 Length: 39  
 Score: 133.00 Matches: 26  
 Percent Similarity: 79.49% Conservatives: 5  
 Best Local Similarity: 66.67% Mismatches: 8  
 Query Match: 9.86% Indels: 0  
 DB: 2 Gaps: 0

## US-09-732-091-3 (1-759) x F64718 (1-39)

QY 1 ATGCGATACAAATATGATAGAGACTTGGAAATTTTAAAGCAATTCAGTATTTATTA 60  
 |||||  
 Db 1 MetAlaTyrArgTyrAspSerAspLeuGluPheLeuLysArgLeuSerSerAspLeu 20  
 |||||  
 QY 61 TTGATTTGTTTTCAGGTGCTTGTGTTTAAAGACGCGCGGAAACACACAAATGAA 117  
 |||||  
 Db 21 LysAspLeuPheAspAlaLeuValTyrAspGluAspGlyThrLeuArgMetAsnGlu 39  
 |||||

RESULT 11  
 D90630  
 hypothetical protein ECs0011 [imported] - Escherichia coli (strain O157:H7, substrain  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: D90630  
 R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: D90630  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-206 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA033435.1; PID:g13359468; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs0011

C;Superfamily: Escherichia coli heat shock protein htga

Alignment Scores: 0.0104 Length: 206  
Pred. No.: 127.50 Matches: 63  
Score: 127.50 Matches: 63  
Percent Similarity: 38.84% Conservative: 24  
Best Local Similarity: 28.12% Mismatches: 96  
Query Match: 9.46% Indels: 41  
DB: 2 Gaps: 7

US-09-732-091-3 (1-759) x D90630 (1-206)

```

Qy 629 ACGCTGTAATGATCCAGCCACAGGACCTGTTAAAGCTCAGAGTCTTGTAGACACC 570
Db 5 ThrProPheSerAlaAlaProThrGlyProProSer----- 16
Qy 569 TGAATGCCCGCA---AGCGATAAACCCAGCCCTAGAAATGGTTTTCGACCGCATTCGCA 513
Db 17 -----ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 34
Qy 512 ACAATGACAGCTAATGATAAGATTAA---CCCCCATTTTAAACAGCGTTAAA 459
Db 35 LysMetArgValSerTrpLeuGluSerArgCysAspThrProPheAlaAsnAsn----- 52
Qy 458 GTCCCGCGCTTAAAGCTTGTCTGTTTAAATGTCGGTGTCTTTTATGGATAATTCATCG 399
Db 53 -----LeuSerPheIleSerSerGlySerSerSer 62
Qy 398 CACATTTCTTTCACT-----TCCTCATCATCC 372
Db 63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuSerSerSer 82
Qy 371 ATTCTTCCAACTCTTCTTAAGATTTTAGAAGCATGTTTGTTCATTAAGTCGTT 312
Db 83 IlePhePheGlnValLeuArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102
Qy 311 TATAGACTCTCGCTTAAATGAACTCGCAAACTATTGCTCCCATAGTATTGCAAC 192
Db 123 TyrAsnLeuProCysProArgSerLeuLeuAlaLeuSerProProLysCysCysAsn 142
Qy 191 TCTTCAGGATCTTCTGCGTATTAGCGTAATCATGCCATGC-----CTT 144
Db 163 AsnSerSerLeuArgIleSerThrLeuGluArgArgValLeuProPheSerLeuTrpVal 182
Qy 83 ACAAGCAGCTCAAAACAAATCCAAATCCAAATGCTTTTAAATAATCCAAAG 24
Db 183 SerSerArgAlaLysPheAlaAsnCysSer---SerLeuGlnCysTrpArgLysSerArg 201
Qy 23 TCTCTATCATAT 12
Db 202 SerGluSerPhe 205

```

RESULT 12

D85481  
Hypothetical protein htga [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: D85481  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamocis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; UID:21074935; PMID:11206551  
A;Accession: D85481  
A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-206 <STO>  
A;Cross-references: GB:AE005174; NID:gl2512690; PID:AA054312.1; GSPDB:GN00145; UMGPI:  
A;Experimental source: strain O157:H7, substrain EDL933

C;Gene: htga  
C;Superfamily: Escherichia coli heat shock protein htga

Alignment Scores: 0.0104 Length: 206  
Pred. No.: 127.50 Matches: 63  
Score: 127.50 Matches: 63  
Percent Similarity: 38.84% Conservative: 24  
Best Local Similarity: 28.12% Mismatches: 96  
Query Match: 9.46% Indels: 41  
DB: 2 Gaps: 7

US-09-732-091-3 (1-759) x D85481 (1-206)

```

Qy 629 ACGCTGTAATGATCCAGCCACAGGACCTGTTAAAGCTCAGAGTCTTGTAGACACC 570
Db 5 ThrProPheSerAlaAlaProThrGlyProProSer----- 16
Qy 569 TGAATGCCCGCA---AGCGATAAACCCAGCCCTAGAAATGGTTTTCGACCGCATTCGCA 513
Db 17 -----ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 34
Qy 512 ACAATGACAGCTAATGATAAGATTAA---CCCCCATTTTAAACAGCGTTAAA 459
Db 35 LysMetArgValSerTrpLeuGluSerArgCysAspThrProPheAlaAsnAsn----- 52
Qy 458 GTCCCGCGCTTAAAGCTTGTCTGTTTAAATGTCGGTGTCTTTTATGGATAATTCATCG 399
Db 53 -----LeuSerPheIleSerSerGlySerSerSer 62
Qy 398 CACATTTCTTTCACT-----TCCTCATCATCC 372
Db 63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuSerSerSer 82
Qy 371 ATTCTTCCAACTCTTCTTAAGATTTTAGAAGCATGTTTGTTCATTAAGTCGTT 312
Db 83 IlePhePheGlnValLeuArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102
Qy 311 TATAGACTCTCGCTTAAATGAACTCGCAAACTATTGCTCCCATAGTATTGCAAC 192
Db 123 TyrAsnLeuProCysProArgSerLeuLeuAlaLeuSerProProLysCysCysAsn 142
Qy 191 TCTTCAGGATCTTCTGCGTATTAGCGTAATCATGCCATGC-----CTT 144
Db 143 SerProAlaIleSerCysGlnLeuArgCysCysSerGlyCysProSerIleAspLeu 162
Qy 143 TTGTATCTATGAGCTGTCAGTCTTTTCAITGTGCTTTTTCGCGCTTTTACCAAAA 84
Db 163 AsnSerSerLeuArgIleSerThrLeuGluArgArgValLeuProPheSerLeuTrpVal 182
Qy 83 ACAAGCAGCTCAAAACAAATCCAAATCCAAATGCTTTTAAATAATCCAAAG 24
Db 183 SerSerArgAlaLysPheAlaAsnCysSer---SerLeuGlnCysTrpArgLysSerArg 201
Qy 23 TCTCTATCATAT 12
Db 202 SerGluSerPhe 205

```

RESULT 13

A40623  
heat shock protein htga - Escherichia coli (strain K-12)  
N;Alternate names: heat shock protein Y  
C;Species: Escherichia coli  
C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 01-Mar-2002  
C;Accession: A40623; A56688; D64721; S28460

R.Missiakas, D.; Georgopoulos, C.; Raina, S.  
 J. Bacteriol. 175, 2613-2624, 1993  
 A:Title: The *Escherichia coli* heat shock gene *htpY*: mutational analysis, cloning, sequen  
 A:Reference number: A40623; MUID:93239687; PMID:8478327  
 A:Accession: A40623  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-196 <MIS>  
 A:Note: sequence extracted from NCBI backbone (NCBI:130422, NCBI:130423)  
 R.James, R.; Dean, D.O.; Debbage, J.  
 DNA Seq. 3, 327-332, 1993  
 A:Title: Five open reading frames upstream of the *dnxK* gene of *Escherichia coli*.  
 A:Reference number: A56688; MUID:94003405; PMID:8400364  
 A:Accession: A56688  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 36-196 <JAM>  
 A:CROSS-references: EMBL:X67700; NID:G41754; PIDN:CAA47932.1; PID:G41757  
 R.Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Co  
 .A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: D64721  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-196 <BLAT>  
 A:CROSS-references: GB:AE000112; GB:U00096; NID:G1786192; PIDN:AACT73123.1; PID:G1786194  
 A:Experimental source: strain K-12, substrain MG1655  
 A:Note: it is uncertain whether Met-1 (CTG) or Met-36 is the initiator  
 C:Genetics:  
 A:Gene: *htgA*, *htpY*  
 A:Map position: 0 min  
 A:Start codon: CTG  
 C:Superfamily: *Escherichia coli* heat shock protein *htgA*  
 C:Keywords: heat shock; stress-induced protein

Alignment Scores:  
 Pred. No.: 0.0295 Length: 196  
 Score: 122.00 Matches: 51  
 Percent Similarity: 38.41% Conservative: 12  
 Best Local Similarity: 31.10% Mismatches: 65  
 Query Match: 9.05% Indels: 36  
 DB: 2 Gaps: 5

US-09-732-091-3 (1-759) x A40623 (1-196)

QY 629 AGCCCTGTAATGATCCAGCCACAGCAGCTGTTAAAGCTCAGAGTCTTGTAAGCACC 570  
 Db 5 ThrProPheSerAlaAlaProThrGlyProProSer----- 16  
 QY 569 TGATGCCCGCA---AGCGATAAACCGCCCTAGAAATGTTTTCGACCGCATTCGCA 513  
 Db 17 -----ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaItpValArg 34  
 QY 512 ACATGACAGCTAATGTAAGATTAAA-----CCCCCATTTTAAACGCTTAAA 459  
 Db 35 LysMetArgValSerTrpLeuGluSerLysCysAspThrProPheAlaAsn----- 52  
 QY 458 GTCGCGCGCTTAAGGCTTGCTGTTAAATCTCCGCTGTTTATGATATTAATTCATCG 399  
 Db 53 -----LeuSerPheIleSerSerGlySerSerSer 62  
 QY 398 CACATTTCTTTCACT-----TCTTCATCATCC 372  
 Db 63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 82  
 QY 371 ATTCTTCCAACTTCTTCTAGATTAGAAAGCATGTTTTCGTTCAATTAAGTCGT 312  
 Db 83 IlePhePheGlnValLeuArgArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102  
 QY 311 TCAGTTTCTTGTTGTAATGACCTTTAAATTTATCGCACATCGCATATAAAATCTCTTG 252  
 Db 311 ----- 196

Db 103 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 122  
 QY 251 TATAAGACTCTTTCGGCTTTAATGAAACTCGCAAACTATTGCTCCCATAGTATTGCAAC 192  
 Db 123 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 142  
 QY 191 TCTTCAGCGATT 180  
 Db 143 SerProAlaIle 146  
 RESULT 14  
 C89910  
 hypothetical protein SALL181 [imported] - *Staphylococcus aureus* (strain N315)  
 C:Species: *Staphylococcus aureus*  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: C89910  
 R;Kuroda, M.; Ohba, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: C89910  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1009 <KUR>  
 A:CROSS-references: GB:BA000018; PID:G13701144; PIDN:BAB42439.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SALL181

Alignment Scores:  
 Pred. No.: 2.56 Length: 1009  
 Score: 98.50 Matches: 66  
 Percent Similarity: 37.88% Conservative: 45  
 Best Local Similarity: 22.53% Mismatches: 83  
 Query Match: 7.30% Indels: 99  
 DB: 2 Gaps: 16

US-09-732-091-3 (-759) x C89910 (1-1009)

QY 13 TATGATAGACATTCGAAATTT-----TTAAGCAATTGGAA 48  
 Db 708 TyrAsnAsnAspLeuGluLeuAsnGluSerLeuGluSerLeuMetSerArgLeuAsn 727  
 QY 49 TCTAGTGATTATTCGATTGTTTGAGGTGCTTTGTTTGGTAAAGACGCGCAAAAGA 108  
 Db 728 LeuThrAspAsnAspIleAsnGluIleAlaItpArgGlyGluGluGluGlu-- 746  
 QY 109 CACATGAAAACTGACAGCTCCATAGATACAAAGGCATGCGATGATTACGCTAAA 168  
 Db 747 -----LeuGluGlnLysArg-----AspThrThrLysLys 756  
 QY 169 TACGCAGAAAGAAATCGCTGAAGATTCGCAATACTATGGGAGCAATAGTTTTCGAGTTTC 228  
 Db 757 ArgCysHisGluPheGluMetGluIleAla-----ArgLeuGluSerLeu 771  
 QY 229 ATTAAGGCGAGAGAGTCTTTATACAAAGAGATTTTATGCGATGTGCGGATAATTAAG 288  
 Db 772 ThrLysAsp-----LysGluLeuLeuAspSer-----AspLysLeuLys 784  
 QY 289 GTCAATTACAAAGAAACTGAA-----ACGACTTTTAATTCGAA----- 327  
 Db 785 AspGluTyrGluGlnLysLysGluLysMetAsnThrLeuIleAspGluTyrSerAlaVal 804  
 QY 328 -----CAAAACATGCTTTCTAAATCTTAGAA 354  
 Db 805 HisTyrGlnCysGlnAsnAsnIleAsnLysThrGlnSerIleValSerHisIleAsnTyr 824  
 QY 355 AGAAGTTTGGAGAAATGATGATGAAGAGAGATTTTATGCGATGTGCGGATAATTAAG 414  
 Db 825 LeuAsnGlnGluLeuLysAspGlnGlnGluIlePheGlnLeuAlaGluIleValSerGly 844



C/Accession: AF030360  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001  
A>Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

US-09-732-091-3 (1-759) x AE0360 (1-282)

Qy	55	GATTATATGGATTTGTTTGGAGTGCTGTGTTTTGTTAAAGACGCCGAATAA---AGACAC	111
		:::::	
Dd	7	GluValLeuAspAlaValcIuArgAsnProTyrSerAsnSerGluAspLysThrArgGln	26
			::::
Qy	112	AATGAAAAACTCACCGCTCCCATAGTAACAAGAAGCATGGCGAT---CATTACGCATAA	168
		::::	
Dd	27	ArgAsnAsrValMetArgSerIleMetTyrAsnLysGlnAsnGluLeuAspGlnValLys	46
			::::
Qy	169	TACGCAGAAAAGAACGCTCAACAGTTGCATTAATCTATGGCAGCAATAGTTTTTCGCAGT---	225
Dd	47	-----AsnSerIleAlaSerLys	52
Qy	226	TTCATTAAAGCGCAGGAGCTCTTATACAAGAGAGATTTTATGCGATGTGTCGCAATAATTA	285
			:::
Dd	53	PheIleLysTrpSer-----LeuIleLysAspGlnPheIleSerGluValThrArgTyr	70

Search completed: July 5, 2004, 03:08:17  
Job time : 27.5 secs

## RESULT 15

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 5, 2004, 02:52:17 ; Search time 14.5 Seconds

(without alignments)

5451.207 Million cell updates/sec

Title: US-09-732-091-3

Perfect score: 1349  
Sequence: 1 atgcatacaaatatgatag.....cgttcgaatagaatccatt 759

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=framer\_n2p.model -DEV=xlh  
-Q=/cgn2.1/USPTO spoil/US09732091/runat 02072004 182406 19007/app query.fasta 1.903  
-DB=SwissProt 42 -OFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -NCPEN=2000000000  
-USER=US09732091 @CGN 1.1.12 @runat 02072004 182406 19007 -NCPEN=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1279	94.8	253	1 YF88_HELPY	Q26107 helicobacte
2	1270	94.1	253	1 YF88_HELPY	Q92124 helicobacte
3	637	47.2	209	1 YF87_HELPY	Q26106 helicobacte
4	630	46.7	209	1 YF87_HELPY	Q92125 helicobacte
5	265	19.6	237	1 YAAW_ECOLI	P75617 escherichia
6	262	19.4	237	1 YAAW_ECO57	P58316 escherichia
7	122	9.1	196	1 HTGA_ECOLI	P28697 escherichia
8	105	7.8	726	1 HSGA_BRARE	Q90474 brachydanio
9	97	7.2	916	1 SECA_TREPA	O83394 treponema p
10	96.5	7.2	373	1 B10F_AQUAE	O66875 aquifex aeo
11	95.5	7.1	344	1 ABIC_LACLA	Q01457 lactococcus
12	95	7.0	517	1 YC09_CAMJE	Q9pn86 campylobact
13	95	7.0	3214	1 BPAL_HUMAN	Q03001 homo sapien
14	94.5	7.0	1433	1 REST_CHICK	Q42184 gallus gall
15	93.5	6.9	802	1 PAC_BACME	Q60136 bacillus me
16	92.5	6.9	802	1 PAC_ARTVI	P31956 arthroabact
17	92.5	6.9	892	1 RCQ1_CABEJ	P46064 caenorhabdi
18	91.5	6.8	399	1 RPSD_THEME	P77994 thermotoga

19	91.5	6.8	472	1	GATB_METJA
20	91.5	6.8	719	1	HS9A_HORSE
21	91.5	6.8	1769	1	YJK9_YEAST
22	91	6.7	284	1	TPM1_CIOIN
23	90	6.7	404	1	LA_BOVIN
24	89.5	6.6	280	1	Y7ES_CHLPN
25	89.5	6.6	633	1	GATE_SULSO
26	89.5	6.6	869	1	AMPN_HABIN
27	89	6.6	386	1	NESG_HUMAN
28	88.5	6.6	732	1	HS9A_CRIGR
29	88	6.5	371	1	RPSD_STRMU
30	88	6.5	670	1	SSM4_SCHPO
31	87.5	6.5	728	1	HS9A_CHICK
32	87.5	6.5	731	1	HS9A_HUMAN
33	87.5	6.5	732	1	HS9A_MOUSE
34	87.5	6.5	732	1	HS9A_PIG
35	87	6.4	607	1	HTPG_FUSNN
36	86.5	6.4	619	1	PRIA_HELPY
37	86.5	6.4	686	1	HMCT_HRLPJ
38	86.5	6.4	1076	1	CARB_IARCFU
39	86.5	6.4	1102	1	CARB_STRCO
40	86	6.4	313	1	OSDI_HUMAN
41	86	6.4	619	1	PRIA_HELPY
42	86	6.4	633	1	YN40_ARCFU
43	86	6.4	1164	1	KELI_YEAST
44	86	6.4	1169	1	EX5B_BORBU
45	85.5	6.3	411	1	HFLK_BUCAP

#### ALIGNMENTS

RESULT 1	YF88_HELPY	STANDARD	PRT	253 AA
ID	YF88_HELPY			
AC	Q26107			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein HP1588			
GN	HP1588			
OS	Helicobacter pylori (Campylobacter pylori)			
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;			
OC	Helicobacteraceae; Helicobacter			
OX	NCBI_TaxID=210;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=26695 / ATCC 700392;			
RX	MEDLINE=97394467; PubMed=9252185;			
RA	Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,			
RA	Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,			
RA	Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,			
RA	Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,			
RA	McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,			
RA	Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,			
RA	Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,			
RA	Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,			
RA	Venter J.C.;			
RT	*The complete genome sequence of the gastric pathogen Helicobacter			
RT	pylori.*;			
CC	Nature 388:539-547(1997).			
CC	-!- SIMILARITY: Belongs to the UPF0174 family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; A800656; AAD08627.1; -			
DR	PIR; D64718; D64718.			
DR	TIGR; HP1588; -			

```
DR InterPro; IPR005367; UPF0174.
DR Pfam; PF03667; UPF0174; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 28417 MW; 00E15A38B1A2036A CRC64;

Alignment Scores:
Pred. No.: 7,56e-96 Length: 253
Score: 1279.00 Matches: 253
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 94.81% Indels: 0
DB: 1 Gaps: 0

US-09-732-091-3 (1-759) x YF88_HELPJ (1-253)

QY 1 ATGGCATACAAATATGATAGAGACTTGGAAATTTTAAAGCAATGGATCTAGTATTGA 60
DB 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20
QY 61 TTGGATTGTTGAGGTGCTGTTTGGTAAAGACGGCGGAAAGACACAAATGAAAAA 120
DB 21 LeuAspLeuPheGluValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40
QY 121 CTGACCGCTCCATAGATCAAAAGCATGGCGATGATTAACGTAATACGCGAAGA 180
DB 41 LeuThrSerSerLeuGluLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60
QY 181 ATGGCTGAAGAGTTCGATGATGAGAGAGTAAAGAGTAAAGAGTAAAGAGTAAAG 240
DB 61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheLysGlyGlu 80
QY 241 GGAGTCTTATACAAAGAGATTTTATGCGATGTCGCGATGATTAAGAGTCAATTA 300
DB 81 GlyValLeuTyrLysGluLysGlnLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100
QY 301 AAGAAAACTGAACGACTTTAAATGAAACAAACATGCTTTCTAAATCTTGAAGAAGT 360
DB 101 LysLysThrGluThrThrLeuLysGlnLeuMetLeuSerLysLeuGluArgSer 120
QY 361 TTGGAAGAAATGATGATGAGAGAGTAAAGAGTAAAGAGTAAAGAGTAAAGAGTAA 420
DB 121 LeuGluGluMetAspAspGluValLysGluMetCysAspGluLeuSerLysLeu 140

Alignment Scores:
Pred. No.: 4,03e-95 Length: 253
Score: 1270.00 Matches: 251
Percent Similarity: 99.60% Conservativeness: 1
Best Local Similarity: 99.21% Mismatches: 1
Query Match: 94.14% Indels: 0
DB: 1 Gaps: 0

US-09-732-091-3 (1-759) x YF88_HELPJ (1-253)

QY 1 ATGGCATACAAATATGATAGAGACTTGGAAATTTTAAAGCAATGGATCTAGTATTGA 60
DB 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20
QY 61 TTGGATTGTTGAGGTGCTGTTTGGTAAAGACGGCGGAAAGACACAAATGAAAAA 120
DB 21 LeuAspLeuPheGluValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40
QY 121 CTGACCGCTCCATAGATCAAAAGCATGGCGATGATTAACGTAATACGCGAAGA 180
DB 41 LeuThrSerSerLeuGluLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60
QY 181 ATGGCTGAAGAGTTCGATGATGAGAGAGTAAAGAGTAAAGAGTAAAGAGTAAAG 240
DB 61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheLysGlyGlu 80
QY 241 GGAGTCTTATACAAAGAGATTTTATGCGATGTCGCGATGATTAAGAGTCAATTA 300
DB 81 GlyValLeuTyrLysGluLysGlnLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100
QY 301 AAGAAAACTGAACGACTTTAAATGAAACAAACATGCTTTCTAAATCTTGAAGAAGT 360
DB 101 LysLysThrGluThrThrLeuLysGlnLeuMetLeuSerLysLeuGluArgSer 120
QY 361 TTGGAAGAAATGATGATGAGAGAGTAAAGAGTAAAGAGTAAAGAGTAAAGAGTAA 420
DB 121 LeuGluGluMetAspAspGluValLysGluMetCysAspGluLeuSerLysLeu 140

DR EMBL; AE001571; AAD07073.1; -.
DR PIR; B71800.
DR InterPro; IPR005367; UPF0174.
DR Pfam; PF03667; UPF0174; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 28475 MW; 127158B2B1A2036A CRC64;

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
```

QY 421 ACCGACAAATTAAACAGACAGACCTTAAGCGCGCGACTTAAACGCTGTTTAAATGGG 480  
 Db 141 ThrAspAsnLeuAsnArgGlcAlaLeuSerAlaAlaThrLeuThrLeuPheLeuMetGly 160  
 QY 481 GGTGTTTAAATCTTAACTAGCTGCTCATTTGCGGATGCGTGGCAAAACCATCTTA 540  
 Db 161 GlyPheLeuSerGlnLeuAlaValileValAlaAsnAlaValAlaLeuThrLeuLeu 180  
 QY 541 GGGCGTGGTTTATCGTTCGGCGCAATCAGAGTGTCTTAAAGAACTCTGAGCTTTTAA 600  
 Db 181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200  
 QY 601 GGTCTGTTGCTGATCATTAACAGCGTATGAGCAGCGATGATGAGCGCGCGCT 660  
 Db 201 GlyProValGlyTyrPheLeuThrGlyValThrAlaIleAspAlaAlaGlyProAla 220  
 QY 661 TATAGGTAAACCATACCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 Db 221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuGlnThrGlnGln 240  
 QY 721 GCCAATGAGATAGAGTCTGCTGCAATAGATCCATT 759  
 Db 241 AlaAsnGluAspLeuSerLeuGlnIleGluSerVal 253  
 RESULT 3  
 YF87\_HELPJ  
 ID YF87\_HELPJ STANDARD; PRT; 209 AA.  
 AC 026106;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein HP1587.  
 GN HP1587.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RA "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori."  
 RL Nature 388:539-547(1997).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Baloch A.;  
 RL Unpublished observations (OCT-2001).  
 CC -!- SIMILARITY: Belongs to the UPF0174 family.  
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to a  
 CC frameshift in position 160.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE000656; AAD08626.1; ALT\_FRAME.  
 DR TIGR; HP1587; -  
 DR InterPro; IPR005367; UPF0174.  
 DR Pfam; PF03667; UPF0174; 1.

KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 209 AA; 23069 MW; F98D3FB8F3F62323 CRC64;

## Alignment Scores:

Pred. No.: 5,46e-44 Length: 209  
 Score: 637.00 Matches: 123  
 Percent Similarity: 76.47% Conservative: 33  
 Best Local Similarity: 60.29% Mismatches: 44  
 Query Match: 47.22% Indels: 4  
 DB: 1 Gaps: 2

US-09-732-091-3 (1-759) x YF87\_HELPJ (1-209)

QY 112 AATGAAATGACCGAGCTCCATGATACAAAGCATGCGATGATACCTAAATAC 171  
 Db 2 AsnGluAspLeuThrAsnSerThrGluTyrLysArgTyrGlyHisAspTyrAlaLysTyr 21  
 QY 172 GCAGAAAGATCGCTGAGAGCTGCAATACATGAGAGCAATAGTTTTCGAGTTTCATT 231  
 Db 22 ProArgArgIleAlaGluGluGlnHisTyrGlyLysSerPheAlaAsnPhePhe 41  
 QY 232 AAAGCGGAGAGCTTATACAAAGAGATTTTATCGGATGTCGATGATAAATAAGGTC 291  
 Db 42 ArgAspGluGlyValLeuTyrLysGlnIleLeuCysAspAlaCysAspHisLeuLysVal 61  
 QY 292 AATTACAAACAGAAACTGAAACGACTTTAATGAAACAAACATGCTTTCTAAATCTTA 351  
 Db 62 AsnTyrAsnGluSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81  
 QY 352 GAAAGAGTTTGGAGAAATGATGATGAGAGAGTGAAGAAATGTCGATCAATATATCC 411  
 Db 82 LysAspSerLeuGluLysSerArgArgGluIleLysGluLeuLysGluLeuGly 101  
 QY 412 ATAAAAACACGACCAATT-----AACACACAGCCCTTAAGCCGCGACTTTA 462  
 Db 102 MetThrAsnIleAspLysValIleGlyLysGlnLysGlnValLeuIleAlaSerThrLeu 121  
 QY 463 ACGCTGTTTAAATGGGGGTTTAAATCTTATCAATAGCTGCTCATTTGCGAATGGG 522  
 Db 122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValSerValAlaAspAla 141  
 QY 523 GTGCAAAACCATCTCTAGGGCGT---GGTTTATGCTTCGGGCAATCAGGTGCTTACA 579  
 Db 142 MetValArgGlnThrLeuGlyHis\*\*\*AlaCysTyrValValGlyLysValAlaLeuLys 161  
 QY 580 AGAACTCTGAGCTTTTAAACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 639  
 Db 162 LysThrLeuGlyValLeuAlaGlyProIleGlyTyrValIleThrGlyAlaLeuValSer 181  
 QY 640 ATTGATATGCGAGCGCGCTTATAGCTTATACCATACCGCATGCTGCTGCTGCTGCT 699  
 Db 182 IleAsnLeuAlaGlyProAlaTyrArgValThrValProAlaCysValLeuIleAlaThr 201  
 QY 700 TTACGCTTAAAA 711  
 Db 202 LeuArgLeuLys 205

## RESULT 4

YF87\_HELPJ  
 ID YF87\_HELPJ STANDARD; PRT; 209 AA.  
 AC 092025;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein JHP1493.  
 GN JHP1493.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 RT gastric pathogen *Helicobacter pylori*."  
 RL Nature 397:176-180(1999).  
 CC -1- SIMILARITY: Belongs to the UPF0174 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE001571; AAD07372.1; -;  
 DR PIR; A71800; A71800.  
 DR InterPro: IPR005367; UPF0174.  
 DR Pfam; PF03667; UPF0174; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 209 AA; 22907 MW; C6C950CDD424CAF CRC64;

Alignment Scores:  
 Pred. No.: 2,01e-43 Length: 209  
 Score: 630.00 Matches: 122  
 Percent Similarity: 75.43% Conservative: 32  
 Best Local Similarity: 59.80% Mismatches: 46  
 Query Match: 46.70% Indels: 4  
 DB: 1 Gaps: 2

US-09-732-091-3 (1-759) x YF87\_HELPJ (1-209)

QY 112 ATGAAACCTGACAGCTCCATAGATACAAAGGATGCGATGATTAGCTAATATAC 171  
 Db 2 AsnGluLeuThrSerLeuThrGluThrGlnArgTyrGlyHisAspTyrAlaLysTyr 21  
 QY 172 GCAGAAAGATCTGAAGAGTGTCAATATCTATGAGGAGCAATAGTTTTCGAGTTTCATT 231  
 Db 22 ProArgGlnLeuAlaGluLeuGlnArgTyrGlyLeuSerPheAlaAsnPhePhe 41  
 QY 232 AARGGCGAAGAGCTTTATCAAGACATTTTATGCGATGTGCGATATTAATTAAGGTC 291  
 Db 42 ArgAspGluGlyValLeuTyrLysGluLeuLeuLeuLeuLeuLeuLeuLeuLeu 61  
 QY 292 AATTACACAGAAACCTGAACGACTTTAATCAACAAACATGCTTTCTTAATCTTA 351  
 Db 62 AsnTyrAsnGluArgSerAlaThrSerLeuLeuGlnGlnGlnGlnGlnGlnGln 81  
 QY 352 GAAAGAGTGTGGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 411  
 Db 82 LysAspSerLeuGluLysMetSerGlyArgGluLeuLeuLeuLeuLeuLeuLeu 101  
 QY 412 ATAAAAACACGGCAATTTA-----AACAGACAGCTTTAAGCGCGGCGACTTTA 462  
 Db 102 MetProAsnLeuAspLysValLeuGlyGluAsnLysGlnValLeuLeuAlaSerValLeu 121  
 QY 463 ACGCTGTTTAAATGGGGGTTTAAATCTTATCAATAGCTGCTGCTGCTGCTGCTGCTG 522  
 Db 122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValAlaValAla 141  
 QY 523 GTCGCAAAACCACTTAGCGCGTGGTTTA---TCGCTTGGCGCAATCAGTGCTTACA 579  
 Db 142 MetValArgGlnThrLeuGlyHisGlyLeuSerSerValValGlyLysValAlaLeuLys 161  
 QY 580 AGAATCTGAGCTTTTAAACAGGTCCTGCTGGCTGGATCATTCAGGCGTATGACAGCG 639  
 Db 162 LysThrLeuAspLeuLeuAlaGlyProIleGlyTrpValIleThrGlyAlaLeuValSer 181  
 QY 640 ATTGATATTCAGGCGCGCTTATAGGTTACCACTACCGGATCATTCGCTGTTGCCACT 699

Db 182 IleAsnLeuAlaGlyProAlaTyrArgValThrValProAlaCysValLeuValAlaThr 201  
 QY 700 TTACGCGCTAAAA 711  
 Db 202 LeuArgLysLys 205  
 RESULT 5  
 YAAW\_ECOLI STANDARD; PRT; 237 AA.  
 ID YAAW\_ECOLI  
 AC P75617; Q47290;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yaaW.  
 GN YAAW OR B0011.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=94003405; PubMed=8400364;  
 RX James R., Dean D.O., Debbage J.;  
 RA \*Five open reading frames upstream of the dnaK gene of *E. coli*.;  
 RT DNA Seq. 3:327-332(1993).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA \*The complete genome sequence of *Escherichia coli* K-12.;"  
 RL Science 277:1453-1474(1997).  
 CC -1- SIMILARITY: Belongs to the UPF0174 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X67700; CAA47934.1; -;  
 DR PIR; C64721; C64721.  
 DR ECGene; EG14340; YaaW.  
 DR InterPro; IPR005367; UPF0174.  
 DR Pfam; PF03667; UPF0174; 1.  
 DR KW Hypothetical protein; Complete proteome.  
 FT CONFLICT 190 190 L -> F (IN REF. 1).  
 SQ SEQUENCE 237 AA; 26665 MW; A36682B2EA116747 CRC64;

Alignment Scores:  
 Pred. No.: 6.16e-14 Length: 237  
 Score: 265.00 Matches: 77  
 Percent Similarity: 48.99% Conservative: 44  
 Best Local Similarity: 31.17% Mismatches: 90  
 Query Match: 19.64% Indels: 36  
 DB: 1 Gaps: 6

US-09-732-091-3 (1-759) x YAAW\_ECOLI (1-237)

QY 7 TACAAATATGATAGAGACTTGGAAATTTTAAAGCAATCTAGTATTATGGAT 66  
 Db 5 TyrLeuAsnSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24  
 QY 67 TTGTTTGGGCTGTGTTTGGTAAAGCGCGGAAACACACATGAAACCTGACC 126  
 Db 25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerValLeuMet 44

```

QY 127 AGCTCCATAGATACAAAGG-----CATGGCGATGATTACGCTAAATACGAGAA 177
Db 45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64
QY 178 AGAATCGCTGAAGAGTTCATATCTATCGGAGCAGTATGTTTTCGAGTTTCATTAAGGC 237
Db 65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84
QY 238 GAAGAGTCTTATACAAAGAGATTTATGCGATGTGCGATTAATAATTAAGTCAATTAC 297
Db 85 HisGlyLysLeuTyArgAlaIleLeuLeuAspValSerLysArgLeuLysLysAla 104
QY 298 AACAGAAAACCTGAACGACTTAAATGCAACAAACATCGCTTCTTAAATCTTAGAAGA 357
Db 105 AspLysGluMetSerThrPheGluIleGluGlnLeuLeuGluGlnPheLeuArgAsn 124
QY 358 AGTTTGAAGAAATGGATGATGAA-----GAA 384
Db 125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144
QY 385 GTGAAGAAATGCGGATGAATATTCATATAAAACACGACATTTAAACAGACAGCC 444
Db 145 ValAsnGluLeuGluGluLeuLeuProLeuMetLysAspLysLeuAlaLysGly 164
QY 445 TTAAGCGCGGCGACTTTAAACGCTGTTTAAATGGGGGTTTTAAATCTTATCAATTAGCT 504
Db 165 ValSer-----HisLeuLeuSerSerGlnLeuThr 174
QY 505 GTCAATTGTTCGAATCGCGTGCAGAAACCCATTTAGGCGGTGGTTTATCGCGGC 564
Db 175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu--LeuArgGly 193
QY 565 AATCAGGTGCTTACAAAGAACTCTGAGCTTTTAAACAGTCTCTGTTGGCTGATCATACA 624
Db 194 -----AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205
QY 625 GGCGTATGCAGCGGATGATGATTCAGGCGCGCTTATAGGTAAACATACCGCGCAGC 684
Db 206 GlyValLysAla-----ValSerGlySerAlaIleArgValThrIleProAlaVal 222
QY 685 ATTGTGGTGGCACTTTACG 705
Db 223 LeuGlnIleAlaCysLeuArg 229

RESULT 6
YAAW_ECO57 STANDARD; PRT; 237 AA.
AC P58316;
DT 16-OCT-2001 (Rel. 40, Created);
DT 16-OCT-2001 (Rel. 40, Last sequence update);
DT 28-FEB-2003 (Rel. 41, Last annotation update);
DE Hypothetical protein yaaW.
GN YAAW OR 200:1 OR ECS0012.
CS Escherichia coli O157:H7.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;

```

```

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: Belongs to the UPF0174 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE005178; AAG54311.1; -.
DR EMBL; AP002550; BAB33434.1; -.
DR PIR; C85481; C85481.
DR PIR; C90630; C90630.
DR InterPro; IPR005367; UPF0174.
DR Pfam; PF03667; UPF0174; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 237 AA; 26681 MW; A25482B2E116759 CRC64;

Alignment Scores:
Pred. No.: 1,08e-13 Length: 237
Score: 262.00 Matches: 76
Percent Similarity: 48.99% Conservative: 45
Best Local Similarity: 30.77% Mismatches: 90
Query Match: 19.42% Indels: 36
DB: 1 Gaps: 6

US-09-732-091-3 (1-759) x YAAW_ECO57 (1-237)
QY 7 TACAATAATGATAGAGCTTGGAATCTTTTAAAGCAATTCGATCTAGTATTATTGGAT 66
Db 5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24
QY 67 TTGTTTGGGTGCTTGTGTTTGGTAAAGCAGCGGAAAAAGACACAAATGAAACCTGACC 126
Db 25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerValLeuMet 44
QY 127 AGCTCCATAGATACAAAGG-----CATGGCGATGATTACGCTAAATACGAGAA 177
Db 45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64
QY 178 AGAATCGCTGAAGAGTTCATATCTATCGGAGCAGTATGTTTTCGAGTTTCATTAAGGC 237
Db 65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84
QY 238 GAAGAGTCTTATACAAAGAGATTTATGCGATGTGCGATTAATAATTAAGTCAATTAC 297
Db 85 HisGlyLysLeuTyArgAlaIleLeuLeuAspValSerLysArgLeuLysLysAla 104
QY 298 AACAGAAAACCTGAACGACTTAAATGCAACAAACATCGCTTCTTAAATCTTAGAAGA 357
Db 105 AspLysGluMetSerThrPheGluIleGluGlnLeuLeuGluGlnPheLeuArgAsn 124
QY 358 AGTTTGAAGAAATGGATGATGAA-----GAA 384
Db 125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144
QY 385 GTGAAGAAATGCGGATGAATATTCATATAAAACACGACATTTAAACAGACAGCC 444
Db 145 ValAsnGluLeuGluGluLeuLeuProLeuMetLysAspLysLeuAlaLysGly 164
QY 445 TTAAGCGCGGCGACTTTAAACGCTGTTTAAATGGGGGTTTTAAATCTTATCAATTAGCT 504
Db 165 ValSer-----HisLeuLeuSerSerGlnLeuThr 174
QY 505 GTCAATTGTTCGAATCGCGTGCAGAAACCCATTTAGGCGGTGGTTTATCGCGGC 564

```

```

Db 175 ArgilleuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193
QY 565 AATCAGGTGCTTACAGAACTCTGAGCTCTTTTACAGGTCTCTGTTGCTGATCATTACA 624
Db 194 -----AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205
QY 625 GGGTATGGACAGGATGATGATGCGCGCGCTTATAGGTAACTACCGGCGTGC 684
Db 206 GlyVallyAla-----ValSerGlySerSerTyrArgValThrIleProAlaVal 222
QY 685 ATTGTGCTGCCATTTACGC 705
Db 223 LeuGlnIleAlaCysLeuArg 229

RESULT 7
HTGA_ECOLI
ID HTGA_ECOLI STANDARD; PRT; 196 AA.
AC P28697;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE Very hypothetical heat shock protein htga (Heat shock protein htpy).
GN HTGA OR HTPY OR B0012.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=34003405; PubMed=8400364;
RA James R., Dean D.O., Debbage J.;
RT "Five open reading frames upstream of the dnaK gene of E. coli.";
RL DNA Seq. 3:327-332(1993).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=KL2 / W110;
RX MEDLINE=93239687; PubMed=8478327;
RA Missiakos D., Georgopoulos C., Raina S.;
RT "The Escherichia coli heat shock gene htpy: mutational analysis,
RT cloning, sequencing, and transcriptional regulation.";
RL J. Bacteriol. 175:2613-2624(1993).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=KL2;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Kagata T., Ishihara A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN=KL2 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1232-1238(1997).
CC -1- FUNCTION: NOT KNOWN. REQUIRED FOR HIGH-TEMPERATURE GROWTH OF
CC E. COLI. IT IS POSSIBLE THAT HTGA PROTEIN SOMEHOW REGULATES EITHER
CC TRANSCRIPTION OF THE RPOH GENE OR THE ACTIVITY OF ITS GENE PRODUCT
CC SIGMA-32. PLAYS A ROLE OPPOSITE THAT OF DNAK, DNAJ, AND GRPE IN
CC TERMS OF HEAT SHOCK REGULATION, ANTAGONIZING THE NEGATIVE OF
CC THESE AND THUS FINE-TUNING THE HEAT SHOCK RESPONSE.
CC -1- INDUCTION: By high temperatures.
CC -1- CAUTION: It is uncertain whether Met-1 or Met-36 is the initiator.
CC -1- CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift.
CC -1- CAUTION: This sequence, according to the EcoSeq database (K. Rudd)
CC is probably not a real protein; therefore this entry will probably

```

```

CC be deleted in future releases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67700; CAA47932.1; ALT INIT.
DR EMBL; L03720; AAA23992.1; -_- INIT.
DR EMBL; L03720; AAA23993.1; ALT INIT.
DR EMBL; D10483; -; NOT_ANNOTATED_CDS.
DR EMBL; AE000112; AAC73123.1; -_-
DR PIR; A40623; A40623.
KW Hypothetical protein; Heat shock.
FT DOMAIN 57 65 POLY-SER.
SQ SEQUENCE 196 AA; 21225 MW; ECA6154160A40993 CRC64;

Alignment Scores:
Pred. NO.: 0.0218 Length: 196
Score: 122.00 Matches: 51
Percent Similarity: 38.41% Conservative: 12
Best Local Similarity: 31.10% Mismatches: 65
Query Match: 9.05% Indels: 36
DB: 1 Gaps: 5

US-09-732-091-3 (1-759) x HTGA_ECOLI (1-196)
QY 629 AGCGCTGTAATGATCCAGCAAGACCTGTTTAAAGCTCAGAGTTCITGTGAAGCACC 570
Db 5 ThrProSerAlaAlaProThrGlyProProSer----- 16
QY 569 TGATTGCCCGCA---AGCGATAAACACCGCCCTAGATGGT---TTCGACCGCATTCGCA 513
Db 17 -----ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTipValArg 34
QY 512 ACAATGACAGCTAATGATTAAGATTAA---CCCCCATTTTAAACAGCGTTTAAA 459
Db 35 LysMetArgValSerTrpLeuGluSerLysCysAspThrProPheAlaAsn--- 52
QY 458 GTGCGCGCGCTTAAGGCTTGCTGTTTAAATGTCGGTGTTT---TATGATAATTTCATCG 399
Db 53 -----LeuSerPheIleSerSerGlySerSerSer 62
QY 398 CACATTTCTTTTCACT-----TCTTCATCATCC 372
Db 63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuSerSerSer 82
QY 371 ATTCTTCCAAACTCTTCTTAAGATTTTAAAGACATGTTTGTTCATTAATTAAGTCGTT 312
Db 83 IlePhePheGlnValLeuArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102
QY 311 TCAGTTTCTTGTGTAATTAATGACCTTAATTTATCGCACATCGCATATAAATCTTTTG 252
Db 103 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 122
QY 251 TATAAGACTCTTCGCTTTTAATGAACTCGCAAAATATTGTCTCCCATGATTATGCAAC 192
Db 123 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 142
QY 191 TCTTCACGCGATT 180
Db 143 SerProAlaIle 146

RESULT 8
HS9A_BRARE
ID HS9A_BRARE STANDARD; PRT; 726 AA.
AC Q50474;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```





RESULT 10	
BIOF AQUAE	
ID_BIOF AQUAE	STANDARD; PRT; 373 AA.
AC	O66875; AC
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	8-amino-7-oxononanoate synthase (EC 2.3.1.47) (AONS) (8-amino-7-
DE	ketopelargonate synthase) (7-keto-8-amino-pelargonic acid synthetase)
DE	(7-KAP synthetase) {L-alanine--pimelyl CoA ligase}.
DE	BIOF OR AQ 626.
GN	BIOF OR AQ 626.
OS	Aquifex aeolicus.
OC	Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX	NCBI Taxid=63363;

SEQUENCE FROM N.A.  
STRAIN=VF5;  
MEDLINE=98196666; PubMed=9537320;  
Decker G, Warren P.V, Gaasterland T., Young W.G., Lenox A.L.,  
Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
"The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus.";  
Nature 392:353-358(1998).

```

RL  Nature 392:353-358(1998).
-!- CATALYTIC ACTIVITY: 6-carboxyhexanoyl-CoA + L-alanine = 8-amino-7-
CC   oxononanoate + CoA + CO(2).
CC   COFACTOR: Pyridoxal phosphate (By similarity).
-!- COPRODUCT: Biotin biosynthesis; first step.
CC   -!- PATHWAY: Biotin biosynthesis; first step.
CC   -!- SIMILARITY: Belongs to class-II of pyridoxal-phosphate-dependent
CC   aminotransferases.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; AE000699; AAC06836.1; -.
DR   PIR; G70355; G70355.
DR   HSRP; P12998; 1850.
DR   InterPro; IPR003408; Ala synthase.
DR   InterPro; IPR004839; Aminotrans_I/II.
DR   InterPro; IPR001917; Aminotrans_II.
DR   InterPro; IPR004723; BioF.
DR   Pfam; PF02490; Ala synthase; 1.
DR   Pfam; PF00155; aminotran_1.2; 1.
DR   TIGRFAMs; TIGR00858; bioF_1.
DR   PROSITE; PS00599; AA TRANSFER_CLASS_2; 1.
KW   Biotin biosynthesis; Transferase; Pyridoxal phosphate;
KW   Complete proteome.
FT   BINDING 223 223 PYRIDOXAL PHOSPHATE (BY SIMILARITY) .
SQ   SEQUENCE 373 AA; 42532 MW; 536B34A5D5F84401 CRC64;
-----
Alignment Scores:
Pred. No.: 2.55 Length: 373
Score: 96.50 Matches: 64

```

Percent Similarity: 36.75% Conservatives: 40  
 Best Local Similarity: 22.61% Mismatches: 102  
 Query Match: 77 Indels: 14  
 DB: 1 Gaps: 14

US-09-732-091-3 (1-759) x BIOF\_AQUAE (1-373)

```

QY 25 TTGAATTTTAAAGCAATGGAATCTAGTATTTATTTGGATTGTTTGGAGTGGCTTGT 84
DB 23 LeuGluGlyValLysAspPheCysSerAsnAspTyrLeu----- 35
QY 85 TTGTGAAGACGGGAAAGACACAAATGAAACCTGACCAGCTCCATAGAAACAAA 144
DB 36 -----GlyLeuArgLysHisProGluValValGluSerIleArgValLeu 51
QY 145 AGGCATGCG-----GATGATTACGCTAAATACGCA 174
DB 52 LysGluAlaGlyLeuGlySerGlyAlaSerGlnLeuValSerGlyTyrThrLysHis 71
QY 175 GAAAGATCGCTGAAGCTTG---CAATACATGGGAGCAATAGTTTGGAGTTTCATT 231
DB 72 ArgGluLeuGluGlyLeuAlaGluPheLysGlyThrGluSerCysValLeuPheGly 91
QY 232 AAAGGC-----GAAGGAGCTCTTATAC 252
DB 92 SerGlyPheLeuAlaAsnValGlyThrIleProAlaLeuValGluGlyAspLeu 110
QY 253 AAAGAGATTTTATGCGATGTGTGGATAAATPAAAGTCAATTAACAAGAAACTGAA 312
DB 111 -----ValLeuSerAsp-----GluLeuAsnHis----- 118
QY 313 ACGACTTTAATGAACAAACATGCTTTCTTAAATCTTAGAAGAGTTTGAAGAAATG 372
DB 119 AlaSerIleIleAspGlyValArgLeuSerLysAlaGlnLysArgValPheLysHis 138
QY 373 GATGATGAAGATGAAGAAATGCGCATGATTTATCCATAAAAAACACGCAATTTA 432
DB 139 AspTyrGluGluLeuGluPheLeuLys-----LysAsnArgLysLysPhe 154
QY 433 AACAGACAGCCTTAAGCGCGCGACTTTAACGCTGTTTAAATGGGGGT----- 483
DB 155 ArgArgValLeuIle-----IleThrAspThrValPheSerMetAspGlyAspValAla 172
QY 484 -----TTAAATCTTATCAATTAGCTGTCAATTCGCAAT 519
DB 173 AspLeuLysArgLeuThrGlnIleCysGluGluTyrAspCysMetLeuTyrIleAspGlu 192
QY 520 CGGTGCGCAAAACCAATCTAGGCGGT---GGTTTATCGCTTGGCGCAATCAG----- 570
DB 193 AlaHisThrThrGlyThrIleGlyLysGlyGlyLeuAspTyrPheGlyIleGluHisLys 212
QY 571 -----GTGCTTACAAAGACTCTGAGCTTTTAAACAGTCTGTTGGCTGGATCAT 621
DB 213 GluTyrIleIleValMetGlyThrLeuSerLysAlaLeuGlySerTyrGlyAlaPheVal 232
QY 622 ACAGCGGTATGACAGCGATTCATATTCAGGCGCGCTTATAGGTAAACCATACCGSCA 681
DB 233 CysGlyThrLysLeuLeuIleAsp-----TyrLeuValAsnLysAlaArg 247
QY 682 TGCATTTGTGGTCCACCTTTACGCCCTAAAAACACAGCAAGCCATCGAGATAAGAGTCG 741
DB 248 SerLeuIlePheSerThr---SerLeuProProSerValCysAlaGlyAlaLysLysAla 266
QY 742 TTGCAATA 750
DB 267 IleGluIle 269

```

RESULT 11

ABIC\_LACLA STANDARD; PRT; 344 AA.  
 ID ABIC\_LACLA  
 AC Q01457;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)

```

DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Abortive phage resistance protein abic.
GN ABIC OR PRF.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG Plasmid pTN20.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB2;
RX MEDLINE=93054365; PubMed=1429469;
RA Durmaz B., Higgins D.L., Klaenhammer T.R.;
RT "Molecular characterization of a second abortive phage resistance
RT gene present in Lactococcus lactis subsp. lactis MB2.";
RL J. Bacteriol. 174:7463-7469 (1992).
CC -!- FUNCTION: PROVIDES RESISTANCE TO BACTERIOPHAGE BY ABORTIVE
CC INFECTION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M95956; AA53569.1; --
CC PIR; A47025; A47025.
CC Plasmid.
SQ SEQUENCE 344 AA; 40128 MW; AA6E41E71B4A003C CRC64;

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
95.50	344		
40.95%	36		
23.81%	51		
7.08%	9		

Best Local Similarity: 7.08%  
 Query Match: 1  
 DB: 9

US-09-732-091-3 (1-759) x ABIC\_LACLA (1-344)

```

QY 19 AGAGACTTGGAAATTTTAAAGCAATGGAATCTAGTAT----- 57
DB 91 LysAsnGlnGluSerLysLysLeuAsnAsnSerAspSerAlaAsnArgGluPheTyr 110
QY 58 -----TTATTCGATTTTTCAGTGTCTTTTGTAAAGACGGCAAAACACACAAT 114
DB 111 SerLeuLeuAspLeuPheLys-----LysGluGlnAsnLysSerGluThr 125
QY 115 GAAAACTGACCGCTCCATAGATAACAAAAGG-----CATGCGCATGAT 159
DB 126 IleLysAlaIleSerPheLeu--TyrLysArgAlaIleAsnAspLysHisGlyAsnSer 144
QY 160 TAGCGTAAA-----TACGACGAAGAATCGCT 186
DB 145 PheIleAsnAspTyrAsnIleAspIleGlyAsnGlyLeuAspPheArgPheAsnLeuPhe 164
QY 187 GAAGACTTGCATATCATCGGAGCATAGTTTTCGAGTTTCATTAAAGCGGAGCAGTC 246
DB 165 GluSerAsnGlnTrpPheSerIleThrTyrAsnSerSerLysTyrLysGlyLysAsn--- 283
QY 247 TTATCAAAAGACGATTTTATCGCATCTGTGCGATAAATTAAGGTCAATTAACACAGAAA 306
DB 184 LeuThrLysGlu-----GlnLysValGluLeuValIleSerArgGln 197
QY 307 ACTGAACGACTTTTAAATGAACAAACATGCTTTCTTAAATCTTAGAAAGAGTTGGAA 366
DB 198 PheAspAspValTyrAsnLysMetSerSerTyrPheLysIlePheHisArgIleLeuLys 217
QY 367 GAATGCGATGATGAAGAGTGAAGAAATGTCGATGAATTATCCATAAAGAACACGGAC 426
DB 218 SerLeuAsnLysArgPheAspGluLysLysLeuAspGluSerAspTyrLysAsnTyrIle 237

```

```

QY 427 AATTAAACAGACAGACGCTTAAGCGCGGCGACTTTAAACGCTGTTTAAATGGGGGTTT 486
Db 238 G.YileuAarghGlnLeuSer----- 245
QY 487 AATCTTCAATAGCTGTCATTGTGCGAATCGGTC--GCAGAACCAATCTAGGG 543
Db 246 ---SerGluGlnLeuValVallLeuLeuAAsnSerLysValLysArgGlyLeuGly 264
QY 544 CGTGGTTTATCGCTGCGGCAATCAGGTG 573
Db 265 LeuGlyLeuLeuGlyThrAsnLeu 274

RESULT 12
YC09_CAMJE STANDARD; PRT; 517 AA.
AC Q9PN86;
DT 16-OCT-2001 (Rel. 40, Created)
DT 18-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0144 protein Cj1209.
GN Cj1209.
OS Campylobacter jejuni.
CC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
CC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felkwell T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -!- SIMILARITY: Belongs to the UPF0144 family.
CC -!- SIMILARITY: Contains 1 HD domain.
CC -!- SIMILARITY: Contains 1 KH domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AL139077; CAB73463.1; -.
DR PIR; C81327; C81327.
DR HAMAP; MF_00335; -.
DR InterPro; IPR006674; HD.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR003607; Met_phosphohydro.
DR InterPro; IPR006675; Unchar_HDIG.
DR Pfam; PF01966; HD; 1.
DR Pfam; PF00013; KH; 1.
DR SMART; SM00471; HDG; 1.
DR SMART; SM00322; KH; 1.
DR TIGRfam; TIGR00277; HDIG; 1.
KW Hypothetical protein; Transmembrane; RNA-binding; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT DOMAIN 206 274 KH.
FT DOMAIN 333 426 HD.
SQ SEQUENCE 517 AA; 58025 MW; 6A799602ED363A25 CRC64;

Alignment Scores:
Pred. No.: 3.41 Length: 517
Score: 95.00 Matches: 54
Percent Similarity: 38.04 Conservative: 43
Best Local Similarity: 21.18% Mismatches: 98
Query Match: 7.04% Indels: 60

```

```

DB: 1 Gaps: 10
US-09-732-091-3 (1-759) x YC09_CAMJE (1-517)
QY 4 GCATACAAATATCATAGACACTTGAATTTTAAAGCAATTTGAATCTAGTATTATTG 63
Db 43 AlaileGluTyrgluAlaGluLeu---lleuleylaspalalylsAsnSerlleleuAsn 61
QY 64 GATTGTTGAGTGCTGTTTGTGTAAGACGCGCAAAAGACACATGAAACACTG 123
Db 62 AlaGluLeuGluVal-----LysLysLysTyrgluGluLysThr 74
QY 124 ACCAGC---TCCATAGATACAAAGGCATGCGCATATTACGTCATATACGACAGAGA 180
Db 75 HisLysPheGlnLysaspPheAsnGlnLysPheaspLysLysGluGlnLys 94
QY 181 ATCGCTGAAGAGTTGCAATATCTATGCGAGCAATAGTTTTCGAGTTTTCATTAAAGGCGAA 240
Db 95 LeuGlnGlnGluGluLysLeuLysGluasp----- 105
QY 241 GGAGTCTTATACAAAGAGATTTTATGC----- 267
Db 106 -----LysGluTyrgluLysSerGlnLysHislleGlnAsnLeuGlnSer 121
QY 268 GATGTGTCGATTAATTAAGGTCATTAACAAGAAAGAACTGAACGACTTTA----- 321
Db 122 AspVal---AspLysLeuLysAsnLysTyrglnGlnLysLeuAspValLeuLyslle 140
QY 322 -----ATTGAACAAACATGCTTTCTTAAATCTTTAGAAAGAGTTG 363
Db 141 LeuGluHisSerThrGlyLeuThrGlnAsnGluAlaLysGluillelleLysLysVal 160
QY 364 GAGAAATGATGATGAAGAGTGAAGAAATGCGCATGATATTCATTAATAAAACACG 423
Db 161 GluGluAsnSerargGluGlnlleAlaHislleValArgLysTyrgluGluAlaLys 180
QY 424 GACAAITTAACACAGAACGCTTAAGCGCGGAGCTTTTAAACGCTGTTTAAATGGGGGT 483
Db 181 AsnGluAlaLysargLysAlaAsnPhellelleAlaGlnAlaThrSerArgPheAlaGly 200
QY 484 TTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATCGGTCGCAAAACCATCTTAGGG 543
Db 201 GluPheAlaAlaGluArgLeuLleAsnValleAsnilleLysAsnAspGluLysGly 220
QY 544 CGTGTTTATCGCTGCGGCAATCAGTCTTACAGAACTCTGAGCTTTTACAGGT 603
Db 221 ArgilleGlyLysGluGlyArgAsnVal-----LysThrLeuGlu----- 234
QY 604 CCGTGTGCTGATCAATACAGCGGATGACAGCGATGATATTCAGGCGCGGCTTAT 663
Db 235 -----MetValLeuGlyValAspillelleleasp----- 244
QY 664 AGGTAACCATACCGGCATGCTGCTGCGCACTTTACGCCTA 708
Db 245 -----AspThrProGlyAlaAlaValSerCysPheAsnLeu 257

RESULT 13
BPAL_HUMAN
ID BPAL_HUMAN STANDARD; PRT; 3214 AA.
AC Q03001; Q12825; Q13266; Q13775; Q96J76; Q96Q75; Q9UGD7;
AC Q9UGD8;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1 isoforms 1/2/3/4/5/8 (230 kDa bullous
DE pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein) (Dystonia
DE musclicorum protein) (Fragment).
GN BPAG1 OR DMH OR DT OR KIAA0728.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]

```



```

DR EMBL; U04850; AAA37185.1; -.
DR PIR; I56317; A40937.
DR Genew; HGNC:1090; BPAG1.
DR MIM; 113810; -.
DR MIM; 600086; -.
DR GO; GO:0005604; C:basement membrane; TAS.
DR GO; GO:0005737; C:cytoplasm; IEP.
DR GO; GO:0005200; P:cytoskeletal organization and biogenesis; TAS.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; TAS.
DR GO; GO:0045104; P:intermediate filament cytoskeleton organiza. .; IEP.
DR InterPro; IPR001589; Actbind actnin.
DR InterPro; IPR001735; Calponin-like.
DR InterPro; IPR001101; Plectin repeat.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00681; Plectin; 5.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00435; spectrin; 3.
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEC; 2.
DR PROSITE; PS00002; SH3; FALSE NEG.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00020; ACTININ 2; FALSE_NEG.
DR PROSITE; PS00021; CH; 2.
KW Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW Structural protein; Cytoskeleton; Cell adhesion; Calcium;
KW Calcium-binding; Alternative splicing.
FT NON_TER 1 1
FT DOMAIN 1 1659
FT DOMAIN 1670 2441
FT DOMAIN 2442 3214
FT DOMAIN 215 439
FT DOMAIN 219 322
FT DOMAIN 335 436
FT DOMAIN 774 851
FT REPEAT 876 959
FT REPEAT 984 1077
FT DOMAIN 1091 1143

Alignment Scores:
Pred. No.: 3,61 Length: 3214
Score: 95,00 Matches: 46
Percent Similarity: 44,75% Conservatve: 35
Best Local Similarity: 25,41% Mismatches: 62
Query Match: 7,04% Indels: 38
DB: 1 Gaps: 9

US-09-732-091-3 (1-759) X BPAL_HUMAN (1-3214)
QY 4 GCATCAAAATATGATAGACTTGGAAATTTTAAAG-----CAATTGGAACTAGTGT 57
Db 2075 SerPheArgAspGluLysGluLeuGluArgLeuGlnLeuCysGlnArgLysSerAspHis 2094
QY 58 TTATTGCTGTTT-----GAGTGCTGTTGTTTGGTAAAGACGGCGAAAAA 105
Db 2095 LeuLysGluGlnPheGluLysSerHisGluGlnLeuGlnAsnLeuLysAlaGluLys 2114
QY 106 AGACAACTGAAATCGCTCGACAGTTG-----CAATCTATGGAGCAATAGTTT 219
Db 2115 GluAsnAspLysLeuGlnArgLeuAsnGluGluLysSerAsnGluCysAla 2134
QY 166 AATACGACGAGAAATCGCTGACAGTTG-----CAATCTATGGAGCAATAGTTT 219
Db 2135 GluMetLeuLysGlnLysValGluGluLeuThrArgGlnAsnAsnGluThrLysLeuMet 2154
QY 220 GCGAGTTTCATTAAGCGGAAGGA-----GTCCTATCAAGAGATTTTATGCGAT 270
Db 2155 MetGlnArgIleGlnAlaGluSerGluAsnIleValLeuGluLysGlnThrIleGlnGln 2174
QY 271 GTGTGCGATAAATTAAGGTC----- 291
Db 2175 ArgCysGluAlaLeuLysIleGlnAlaAspGlyPheLysAspGlnLeuArgSerThrAsn 2194

```

---

```

QY 292 ---AATTACAAAGAAACTGAAACGACTTTTAATTGAAACAAACATGCTTTCTAAATC 348
Db 2195 GluHisLeuHisLysGlnThrLysThr-----GluGlnAspPheGlnArgLysIle 2211
QY 349 -----TTAGAAAGAGTTTGGAAAGAAATCGATGAT-----GAAGAAGTGAAGAAATG 396
Db 2212 LysCysLeuGluAspLeuAlaLysSerGlnAsnLeuValSerGluPheLysGlnLys 2231
QY 397 TGGGATGAATTAATTCATA-----AAAAACACGGCAATTTAAACAGA 438
Db 2232 CysAspGlnGlnAsnIleIleGlnAsnThrLysLysGluValArgAsnLeuAsnAla 2251
QY 439 CAA 441
Db 2252 Glu 2252

RESULT 14
REST CHICK
ID REST CHICK STANDARD; PRT; 1433 AA.
AC 042184; 042228; 057563; 057564;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Restin (Cytoplasmic linker protein-170) (CLIP-170).
GN RSN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9813792; PubMed=9463933;
RA Gripapic L., Volosky J.M., Keller T.C. III;
RT "Cloning and expression of chicken CLIP-170 and restin isoforms.";
RL Gene 206:195-208(1998).
RN [2]
RP SEQUENCE OF 17-1139 FROM N.A. (ISOFORMS 3 AND 4).
RX TISSUE=Pectoralis muscle;
RA Gripapic L., Keller T.C. III;
RT "Identification and expression of two novel CLIP-170/Restin isoforms expressed predominantly in muscle.";
RL Biochim. Biophys. Acta 1405:35-46(1998).
CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE CYTOSKELETON (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=042184-1; Sequence=Displayed;
CC Name=2;
CC IsoId=042184-2; Sequence=VSP_000761;
CC Name=3; Synonyms=CLIP-170(11);
CC IsoId=042184-3; Sequence=VSP_000762, VSP_000763;
CC Name=4; Synonyms=CLIP-170(11+35);
CC IsoId=042184-4; Sequence=VSP_000764;
CC -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AF014012; AAC60344.1; -.
CC EMBL; AF020764; AAC60345.1; -.
CC EMBL; AF045650; AAC03547.1; -.

```

DR	EXML; AF045651; AAC03548.1; -.
DR	InterPro; IPR030938; CAP-Gly.
DR	InterPro; IPR031878; Znf_CCHC.
DR	Pfam; PF01302; CAP GLY; 2.
DR	SMART; SM001343; Znf C2HC; 1.
DR	PROSITE; PS00845; CAP_GLY_1; 2.
DR	PROSITE; PS50245; CAP_GLY_2; 2.
KW	Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
FT	DOMAIN 79 121 CAP-Gly 1.
FT	DOMAIN 144 207 SER-RICH.
FT	DOMAIN 235 277 CAP-Gly 2.
FT	DOMAIN 305 332 SER-RICH.
FT	DOMAIN 351 1353 COILED_COIL (POTENTIAL).
FT	DOMAIN 1414 1427 CCHC-BOX.
FT	VARSPLIC 458 492 Missing (in isoform 2).
FT	VARSPLIC 458 492 /Ftid=VSP_000761.
FT	/TQKLHARKELEQSILFEKTKADLKQRELEDTR -> RK
FT	ROISEDEN (in isoform 3).
FT	/Ftid=vsp_000762.
FT	S -> GGSKVS (in isoform 3).
FT	/Ftid=vsp_000763.
FT	T -> RKRQISEDEN (in isoform 4).
FT	/Ftid=vsp_000764.
FT	K -> R {IN REF. 2; AAC03547}.
FT	E -> V {IN REF. 2; AAC03548}.
SQ	SEQUENCE 1433 AA; 161026 MW; 563LCB68349BE23 CRC64;
Alignment Scores:	
Pred. No.:	3.86 Length: 1433
Score:	Matches: 52
Percent Similarity:	Conservative: 36
Best Local Similarity:	Mismatches: 74
Query Match:	Indels: 10
DB:	Gaps: 51

US-09-732-091-3 (1-759) X REST CHICK (1-1433)

Qy	1	ATGCATCAAAATATGATAGACATTCGGAAATTTTAAAGCAATTCGAACTCTAGTGATTGA	60
Db	1020	MetIleThrLysHisAspAlaAspIleuGlyGlyPheLysGln-----AsnLeu	1035
Qy	61	TTTGATTGTTTGGAGTGCTGTTGTTTGGTAAAGACGGCGAAAAAAGACACAAATGA	120
Db	1036	LeuAspAlaGluGluAlaLeu-----LysAlaAlaGlnIlyLysAsnAspGluLeu	1052
Qy	121	CTGACCGAGTCCATAGATATCAAAAGCATGGCGATGATTACGCTCAATAATACGCAGAAAGA	180
Db	1053	GluThrGlnAlaGluGluLeuLysGlyGlnAlaGluGln--AlaLysAlaAspLysArg	1071
Qy	181	ATCCTCGAAGAGTTGCAATACTATGGGAGCAATAGTTTTCGAGTTTCATTAAAGGCGAA	240
Db	1072	AlaGluGluValLeuGln-----ThrMetGluLysValThrLysGluLys	1086
Qy	241	GGAGTCTTATACAA-----GAGATTTTATCGCATGTGTGCGGATAAATAAAGGTC	291
Db	1087	AspAlaIleHisGlnGluLysIleGluThrLeuAlaSerLeuGluAsnSerArgGlnThr	1106
Qy	292	AATTACACACAGAAACATGAAACACACTTAAATGAACAAAACATGCTTCTTAAATCTTA	351
Db	1107	AsnGluLysLeuGlnAsnGluLeuAspMetLeuLysGlnAsnLeuLysAsnGluGlu	1126
Qy	352	GAA-----AGNAGT---TTGGAGAATATGATGATGAAGAAGTCMAAGAAATCTGC	399
Db	1127	GluLeuThrLysSerLysGluLeuLeuAsnLeuGluAsnIlyLysValGluGluLeuLys	1146
Qy	400	GATGAAATTATCCATAAAAAACACGCACAATTAAACACAGACGCTTAAAGCGCGGCACT	459
Db	1147	LysGluPhe-----GluAlaLeuLysLeuAlaAla	1156
Qy	460	TTAAGCGTGT---TAAATCGGGG---TTAAATCTTATCAATTAGCTGTCAT---GTTGCG	516
Db	1157	AlaGln-----LysSerGlnGlnLeuAlaLeuGlnGluGlu	1169

```

Qy 517 AATCGCGTCCAAAAACCACTTCTAGGCGCTGT-----549
    |||
Db 1170 AsnVallysLeuAlaGluGluLeuGlyArgSerArgAspGluValThrSerHisGlnLys 1189
    |||
Qy 550 -----TTATCGCTTGGCGGCAATCAGGTGCTT 576
    |||
Db 1190 LeuGluGluGluArgSerValLeuAsnAsnGlnLeuLeu 1202
    |||
RESULT 15
PAC_BACME STANDARD; PRT; 802 AA.
AC Q60136; Q9S463;
AT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, last sequence update)
DT 15-MAR-2004 (rel. 43, last annotation update)
DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE (Penicillin G amidohydrolase).
GN PAC OR PGA.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14945;
RA Kang J.H., Kim S.J., Park Y.C., Hwang Y., Yoo O.J., Kim Y.C.;
RT "Nucleotide sequence of the penicillin G acylase gene from
RT Bacillus megaterium and characteristics of the enzyme.";
RL Misaimurthag Hoji 32:215-221(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14945;
RX MEDLINE=95180705; PubMed=7875576;
RA Martin L.M., Prieto A.M., Cortes E., Garcia J.L.;
RT "Cloning and sequencing of the pac gene encoding the penicillin G
RT acylase of Bacillus megaterium ATCC 14945.";
RL FEWS Microbiol. Lett. 125:287-292(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CA4098;
RX PubMed=12114980;
RA Wang S., Huang Y.H., Huang X.D., Li S.Y., Yuan Z.Y.;
RT "High expression of penicillin G acylase gene from Bacillus megaterium
RT in Bacillus subtilis.";
RL Acta Biochim. Biophys. Sin. 31:601-603(1999).
CC -- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-
CC aminopenicillanate.
CC -- COPACTOR: Binds 1 calcium ion per subunit (Potential).
CC -- SUBUNIT: Heterodimer of an alpha chain and a beta chain (By
CC similarity).
CC -- SUBCELLULAR LOCATION: Extracellular (Potential).
CC -- SIMILARITY: Belongs to peptidase family S45.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U07682; AAB41343.1; -.
CC EMBL; Z37542; CAA85774.1; -.
CC EMBL; AF161313; AAD45609.1; -.
CC PIR; S49252; S49252.
CC HSP; P06875; IAUQ.
CC MEROPS; S45.001; -.
CC InterPro; IPR002692; Peptidase S45.
CC Pfam; PF01804; Penicill_amide; 1.
CC Hydrolase; Antibiotic resistance; Zymogen; Calcium-binding; Signal.
KW SIGNAL
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 802 PENICILLIN G ACYLASE ZYMOGEN.
FT CHAIN 25 234 PENICILLIN G ACYLASE ALPHA SUBUNIT.
FT PROPEP 235 265 SPACER PEPTIDE.

```

Db 249 GluAsnPheValGlnThrSerGluGluLeuGlyLeuProLeuLysIleGlySerAsnAla 268  
QY 532 ACCATTCTAGGGCGTGGTTTATCGCTTGGGGCAATCAGGTGCTTACAGAACTCTGAGC 591  
Db 269 AlaIleValGlySerGluLysSerAlaThrGlyAsnAlaLeuLeu----- 283  
QY 592 TTTTAAACAGGTCCT---GTGGCTGGATCAITACAGGCGGTATGGACAGCGATTGATATT 648  
Db 284 ---PheSerGlyProGlnValGlyPheValAlaProGlyPheLeuTyrGluValGlyLeu 302  
QY 649 GCAGGGCGGCTTAT 663  
Db 303 HisAlaProGlyPhe 307

Search completed: July 5, 2004, 03:06:43  
Job time : 20.5 secs

FT CHAIN 266 802 PENICILLIN G ACYLASE BETA SUBUNIT.  
FT ACT SITE 266 266 BY SIMILARITY.  
FT METAL 177 177 CALCIUM (POTENTIAL).  
FT METAL 341 341 CALCIUM (POTENTIAL).  
FT VARIANT 3 3 T -> M (IN STRAIN CA4098).  
FT VARIANT 224 224 D -> E (IN STRAIN CA4098).  
FT VARIANT 232 232 I -> K (IN STRAIN CA4098).  
FT VARIANT 254 254 T -> S (IN STRAIN CA4098).  
FT VARIANT 349 349 A -> T (IN STRAIN CA4098).  
FT VARIANT 470 470 S -> N (IN STRAIN CA4098).  
FT VARIANT 524 524 F -> Y (IN STRAIN CA4098).  
FT VARIANT 569 569 Q -> P (IN STRAIN CA4098).  
FT VARIANT 586 586 I -> A (IN STRAIN CA4098).  
FT VARIANT 657 657 N -> S (IN STRAIN CA4098).  
FT VARIANT 740 740 T -> K (IN STRAIN CA4098).  
FT VARIANT 789 791 NKA -> YKS (IN STRAIN CA4098).  
SQ SEQUENCE 802 AA; 91987 MW; 877CA0564E50DFED CRC64;

Alignment Scores:  
Pred. No.: 4.57 Length: 802  
Score: 93.50 Matches: 62  
Percent Similarity: 35.44% Conservative: 39  
Best Local Similarity: 21.75% Mismatches: 107  
Query Match: 6.93% Indels: 77  
DB: 1 Gaps: 10

US-09-732-091-3 (1-759) x PAC\_BACME (1-802)

QY 10 AATATGATAGACACTGGATTTTAAAGCAATCGAATCTAGTATTGATTGATTG 69  
Db 33 LysValValArgAspAsnPheGlyValProHisLeuTyrAlaLysAsnLysLysAspLeu 52  
QY 70 TTTGAGGTG-----CTTGTGTTTGGTAAAGACGGC-----GAAAAA 105  
Db 53 TyrGluAlaTyrGlyTyrValMetAlaLysAspArgLeuPheGlnLeuGluMetPheArg 72  
QY 106 AGACACAATGAAACACTGACAGCTCCATGATACAAAGGCGATGCGGATGATTACGCT 165  
Db 73 ArgGlyAsnGluGlyThrValSerGluIle-----PheGlyGluAspTyrLeu 88  
QY 166 -----AAATACCGCAGAAAGAAAGTCCCT 185  
Db 89 SerLysAspGluGlnSerArgAspGlyTyrSerAsnLysGluIleLysLysMetIle 108  
QY 187 GAAGAGTTGCACACTATGGAGCAATAGTTTCCGAGTTTCATTAAGGCGAAGAGATC 246  
Db 109 AspGlyLeuAspArgGlnProLysGluLeuIleAlaLysPheAlaGluGlyIleSerArg 128  
QY 247 TTATACAAAGAGATTTTATCGGATGTGTGGATAAATTAAGTCAATTACAAC----- 300  
Db 129 TyrValAsnGluAlaLeuLysAspProAspAspLysLeuSerLysGluPheHisGluTyr 148  
QY 301 -----AAGAAACCTGAAACGACTTTTAATTTGAACAAACATGCTTTCT--- 342  
Db 149 GlnPheLeuProGlnLysTyrThrSerThrAspValValArgValTyrMetValSerMet 168  
QY 343 -----AAATCTTAGAAAGAGTTTG 363  
Db 169 ThrTyrPheMetAspAsnHisGlnGluLeuLysAsnAlaGluIleLeuAlaLysLeuGlu 188  
QY 364 GAAGAAATGGATGATGAGAGAGTGAAGCAATGTCGATGATTAATCCATAAAAAAC--- 420  
Db 189 HisGluTyrGlyThrGluValSerArgLysMetPheAspAspLeuValTyrPlyAsnAsp 208  
QY 421 -----ACGCACAAATTA 432  
Db 209 ProSerAlaProThrSerIleValSerGluGlyLysProLysArgAspSerSerGln 228  
QY 433 AACAGACAAAGCCTTAAGCGCGCGACTTTTAAACGCTGTTTAAATGGGGGGTTTAAA--- 489  
Db 229 SerLeuGlnIleLeuSerSerAlaValIleLysAlaSerGluLysValGlyLysGluArg 248  
QY 490 -----TCATTATCAATTAGCTGCTATTGTTGCGAATGCGGTGCAAAA 531

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 5, 2004, 02:09:47 ; Search time 80.5 Seconds

(without alignments)

5949.772 Million cell updates/sec

Title: US-09-732-091-3

Perfect score: 1349

Sequence: 1 atggcatacaatatgatag.....cgt-gcaaatagaatccatt 759

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2\_1/USPTC\_spool/US09732091/runat\_02072004\_182407\_19014/app.query.fasta\_1.903  
-DB=SPTRMBL\_25 -OPMT=fastan -SUFFIX=n2p.rspt -MINVATC=0.1 -LOOPGL=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFM=pco -NORM=ext -HEAPSIZ=500 -WINLEN=0 -MAXLEN=2000000000  
-USER=US09732091 -CGN\_1\_171 -runat\_02072004\_182407\_19014 -NCPU=6 -ICPJ=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL\_25.\*

1: sp.archaea.\*  
2: sp.bacteria.\*  
3: sp.fungi.\*  
4: sp.human.\*  
5: sp.invertebrate.\*  
6: sp.mammal.\*  
7: sp.mhc.\*  
8: sp.organelle.\*  
9: sp.phage.\*  
10: sp.plant.\*  
11: sp.podent.\*  
12: sp.virus.\*  
13: sp.vertebrate.\*  
14: sp.unclassified.\*  
15: sp.rvirus.\*  
16: sp.bacteriap.\*  
17: sp.archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	265	19.6	237 16	Q8XGV3 salmonella

2	260	19.3	237 16	Q83SR4
3	259	19.2	237 16	Q8FLC7
4	133	9.9	39 16	O26108
5	127.5	9.5	206 16	Q8XA70
6	125	9.3	199 16	Q7UDU2
7	105	7.8	1245 5	O81AN1
8	103.5	7.7	403 16	Q8R924
9	102.5	7.6	276 16	Q8DIH4
10	102	7.6	1590 5	Q8MX30
11	101	7.5	1455 5	Q8IKG8
12	100.5	7.4	726 13	P87397
13	99.5	7.4	556 16	Q7VHM0
14	99.5	7.4	633 16	Q8E2L0
15	99	7.3	202 5	Q8IJZ2
16	99	7.3	222 4	O43633
17	99	7.3	222 11	Q9DB34
18	99	7.3	723 13	Q7Z2U9
19	98.5	7.3	1009 16	Q9UDU0
20	98	7.3	282 16	Q8ZCL9
21	98	7.3	413 9	Q8SDW1
22	98	7.3	1099 16	Q98QT2
23	98	7.3	2359 5	Q8I297
24	98	7.3	2849 5	Q8IHY4
25	97.5	7.2	311 16	Q8PIC9
26	97.5	7.2	728 16	Q9FI94
27	97.5	7.2	792 10	Q9ARQ9
28	97	7.2	346 16	O51329
29	96.5	7.2	1102 16	Q827Q7
30	96.5	7.2	1341 5	O8IEQ5
31	96	7.1	262 5	O61768
32	96	7.1	420 16	Q97MM5
33	96	7.1	4405 5	Q8ILZ2
34	95.5	7.1	944 17	Q8RPEC2
35	94.5	7.0	387 10	Q9FX81
36	94.5	7.0	406 16	Q8RHS7
37	94	7.0	385 2	Q9Z188
38	94	7.0	576 10	Q84LB2
39	93.5	6.9	408 10	Q94JW4
40	93.5	6.9	681 10	Q9XIK2
41	93.5	6.9	701 17	Q973X0
42	93.5	6.9	1413 5	Q8I589
43	93	6.9	355 17	Q97UF7
44	93	6.9	610 16	O51577
45	93	6.9	667 11	O35267

## ALIGNMENTS

RESULT 1  
Q8XGV3 PRELIMINARY; PRT; 237 AA.  
ID Q8XGV3  
AC Q8XGV3;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 21-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Hypothetical protein (Positive regulator for sigma H (Sigma 32)  
DE Promoters, permitting growth at high temperature).  
GN T0010 OR HTGA OR STM0010 OR STY0010.  
OS Salmonella typhi, and  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_taxid=601, 602;  
RN [1]  
RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;  
RC MEDLINE=22531367; PubMed=12644504;  
RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
RN [2]









```

Db 163 AsnSerSerLeuArgIleSerThrLeuGluArgValLeuProPheSerLeuTrpVal 182
QY 83 ACAAGCACTCAACAAATCCAAATAATACATAGATTCOAATTGCTTTAAAAATCCAG 24
Db 183 SerSerArgAlaLysPheAlaAsnCysSer---SerLeuGlnCysTrpArgLysSerArg 201
QY 23 TCTCTATCATAT 12
Db 202 SerGluSerPhe 205

RESULT 6
Q7UDU2 PRELIMINARY; PRT; 199 AA.
AC Q7UDU2:2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Positive regulator for sigma 32 heat shock promoters.
GN HTGA OR S0011.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RA MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Ruygen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RA "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AB016978; AAP15558.1; -.
KW Heat shock.
SQ SEQUENCE 199 AA; 21606 MW; 3598441A948498B4 CRC64;

Alignment Scores:
Pred. No.: 0.101 Length: 199
Score: 125.00 Matches: 63
Percent Similarity: 36.49% Conservative: 18
Best Local Similarity: 28.38% Mismatches: 81
Query Match: 9.27% Indels: 60
DB: 16 Gaps: 8

US-09-732-091-3 (1-759) x Q7UDU2 {1-199}

QY 629 ACGCTGTAAATGATCCAGCCCAACAGCACTGTAAAGAGCTCAGAGTTCTTGTAAAGCACC 570
Db 8 ThrProPheAsnAlaAlaProThrGlyProProSer----- 19

QY 569 TGATTGCCGCA---ACGGATAAACCGCCCTAGAAAGTGTGTTTGGACCGCATCGCA 513
Db 20 -----ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 37

QY 512 ACAATGACAGCTAATGTATAGATTAA---CCGCCATTTTAAACAGCGCTTAA 459
Db 38 LysMetArgValSerTrpLeuGlnSerLysCysAspThrProPheAlaAsnAsn----- 55

QY 458 GTCCGCGCGCTTAAGGCTTGCTGTTTAAATGTCCGTTGTTTATGGATAATTCATCG 399
Db 56 -----LeuSerPheIleSerSerSerSer 65

QY 398 CACATTTCTTCACCT-----TCTTCATCATCC 372
Db 66 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 85

QY 371 ATTTCCTCCAACTCTTCTAAGATTTTAGAACGATGTTTGTTCATTAAGTCGTT 312
Db 86 IlePheGlnValLeuArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 105

QY 311 TCAGTTTCTTGTGTAATTGACCTTTAATTATCGCACATCGCATCAATAAATCTCTTG 252

```

```

Db 106 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 125
QY 251 TATAAGACTCTTCGCGCTTTAATGAAGAACTCGCAAACTATGTCCCATAGTATTGCAAC 192
Db 126 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysAsn 145
QY 191 TCTTCAGCGATTCTTCTCGGTATTAGCGTAATCATCGCCATGC----- 147
Db 146 SerProAlaIleSerCysGlnLeuArgArgCysCysSerGlyCysProSerIleAspLeu 165
QY 146 -----CTTTGTATTCTATG-----GAGCTGCTCAGTTTTCATTGTGCTTT 105
Db 166 AsnSerSerLeuArgIleSerMetLeuGluArgArgValLeuProPheSerLeuTrpVal 185
QY 104 TTTTCGCGCTCTTTACCAAAACAAGCACCTCAACAAATCCAAATAATCAGTATTGCC 45
Db 186 -----SerAsnArgAlaLysPheAla 192
QY 44 AATTCG 39
Db 193 AsnCys 194

RESULT 7
Q8IAN1 PRELIMINARY; PRT; 1245 AA.
AC Q8IAN1:2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF08_0127.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RE SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844507; CAD51332.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1245 AA; 147911 MW; D856486AFDFE4DDF CRC64;

Alignment Scores:
Pred. No.: 3.84 Length: 1245
Score: 105.00 Matches: 35
Percent Similarity: 42.42% Conservative: 21
Best Local Similarity: 26.52% Mismatches: 48
Query Match: 7.78% Indels: 28
DB: 5 Gaps: 5

US-09-732-091-3 (1-759) x Q8IAN1 {1-1245}

QY 100 GAAAAAGACACAAATGAAAGAACTGACAGCTCCATAGATAACAAAAGCATGCGCAT--- 156
Db 542 GluLysLysGluGluLysIleMetLysLysLeu---TyrSerGluHisProAspLeu 560
QY 157 -----GATTACGCTAAATACGACAGAAAGATCGCTGAAGATTGCAATATCTATGGAGC 210
Db 561 ValLysAspLeuMetAsnTyrGluLysGluTyrAlaGluLysArgAsnLeuLeuAsn 580
QY 211 AATAGTTTTCGAGTTTTCATTAAAGCGCAGAGCTTATACAAAGAGATTTTATGCGAT 270
Db 581 IleLysLysArgLysThrIleLysThrLeuAsn---ArgTyrLysAsn----- 595
QY 271 GTGTGCGCAATAATAAGGTCAATTTACACAAAGAAACTGAAACGACTTTTAATTGAACAA 330
Db 596 -----GluGlnLeuLysLysGluLysSerIleGlnSerAsnLysGluGluLys 613
QY 331 AACATGCTTTCTAAATCTTAGAAGAGATTG----- 363
Db 614 AsnMetLeuLysLysIleGluLysThrAlaLeuThrGluGlnThrAsnGluAspAsn 633

```

```

QY 364 -----GAAGAAATGGATGATGAGAGAGTGAAGAAATGTC 399
Db 634 AsnMetMetIleHisAspAspAsnAspAspAspAspAspAspMetSerGluAspThr 653
QY 400 GATGAATTATCCATAAAACACGCAATTTAAAC 435
Db 654 AspGluIleSerSerLysAspHisAspAspMetAsn 665

RESULT 8
Q8R9Z4
ID Q8R9Z4 PRELIMINARY; PRT; 403 AA.
AC Q8R9Z4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Hypothetical protein TTE1436.
GN TTE1436.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMEL; A8013102; AAM2458.1;
DR GO; GO:0003424; C:Flagellar hook (sensu Bacteria); IEA.
DR GO; GO:0003774; F:Motor activity; IEA.
DR GO; GO:0009296; P:Flagella biogenesis; IEA.
DR InterPro; IPR001635; Flag hook.
DR Pfam; PF02120; Flg hook; I.
DR PRINTS; PR01007; FEGHOOKFLIK.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 403 AA; 46116 MW; 50C22B9FB51A56F CRC64;

Alignment Scores:
Pred. No.: 5.08 Length: 403
Score: 103.50 Matches: 53
Percent Similarity: 35.14% Conservative: 47
Best Local Similarity: 20.13% Mismatches: 98
Query Match: 7.67% Indels: 105
DB: 16 Gaps: 12

US-09-732-091-3 (1-759) x Q8R9Z4 (1-403)

QY 28 GAATTTTAAAGCAATGGATCTAGTATTTAGTATTTGTTGAGGTGTTGTTT 87
Db 61 GluPheLeuSerGlnLysAsnGlyLeuLysAsnValGluIlePheGln----- 76
QY 88 GGTAAAGACGGGAAAGAACACACAAAGAAAA----- 120
Db 77 -----GluLysValLysGluAspLysPheMetGluAspLeuAsnAspPhe 92
QY 121 -----CTGACGAGTCCATAGAA-----TACAAAAGGCAT 150
Db 93 IleProAlaLeuMetGlnThrIleThrLeuSerLeuGluThrIleSerSerGluLysGly 112
QY 151 GCGGATGATTACCTTAATACGACAGAAAGATCGCTGACAGGTGCAATCTATCGGAGC 210
Db 113 AlaSerAspPheGluLysValArgGluLysLeuGluValAlaLeuGln----- 128
QY 211 AATAGTTTTGCGAGTTTCATTAAGCGCGAGGAGTCTTATACAAAGAGATTTTATCGCAT 270
Db 129 -----GlyPheIleLysGluArgAsnPheThrPheLysGluIleAlaLysLys 144
QY 271 GTGTGCGATAAATTAAGGTCAATTACACAGAAAGAACTGAAACGACTTTAATTGAACAA 330
Db 271 GTGTGCGATAAATTAAGGTCAATTACACAGAAAGAACTGAAACGACTTTAATTGAACAA 330

```

```

Db 145 ILSerAspPheLeuLysGluAsnPheAsnIleGluLeuSerProGluValIleGluArg 164
QY 331 AACATG---CTTTCTAAATCTTAGAAGAAAGTTTGGAGAAATGGATGATGAAGAAAGTG 387
Db 165 HisIleLysLeuAlaLysValLysAspLeuAspLysProPheLeuGlnAspLeuAsnGln 184
QY 388 AAGAAATGCGCATGAATATCCATAAAACACGAC-----AATTTA 432
Db 185 LysAspPheAlaGluGluAsnValGlnLysAsnGlnAspLysThrSerGlnLeuLysIle 204
QY 433 AACAGACAGCCCTTAAGCGCGCG----- 456
Db 205 AspLysGluAlaPheIleAlaLysGluAlaLysGluLysThrGluLysLysSer 224
QY 457 -----ACTTTAACGCTGTTTAAATAGGGGTTTAAATCTTATCAATTA 501
Db 225 PheAspValLysGlnGluPheValPhePheLysAsnGluGly-----LysProVal 241
QY 502 GCTGTCATTTGCGAATGCGTCCGCAAAACCATTCAGGGCGTGGTTCGTCGATCAT 621
Db 242 SerAsnLeuThrTyrAsnSerIleLysLysSer----- 252
QY 562 GGCATTCAGGTCTTACAGAACTCTGAGCTTTTAAACAGTCTCTGTTGGCTGGATCAT 621
Db 253 ---AsnAspProValAspArgLeuPheArg-----GlnIleVal 264
QY 622 ACAGGCGTATGAGACGATTTGATTTGAGGCGCGGCTTATAGGGTAACCATACCGGCA 681
Db 265 AspAsnValPheValAlaLysGluLysGlyAlaSerSerValThrValAsnLeuLysPro 284
QY 682 TGCATT-----GTGGTGGCCACT 699
Db 285 GluIleLeuGlyLysLeuGlnIleSerLeuLysSerIleAspGlyAsnIleValAlaThr 304
QY 700 TTA-----CGCTTAAACACACAGCAA 720
Db 305 IleValThrGluSerGluLysThrLysHisGlnIleGluSerAsnLeuSerLeuGln 324
QY 721 GCCAATGGAGATGAAGATCGTTGCAATAGATCCATT 759
Db 325 AlaGlnLeuAspLeuLysGlyIleLysIleGluSerVal 337

RESULT 9
Q8DIH4
ID Q8DIH4 PRELIMINARY; PRT; 276 AA.
AC Q8DIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tll1614 protein.
GN Tll1614.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabara S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005374; BACC9166.1; -.
KW Complete proteome.
SQ SEQUENCE 276 AA; 31282 MW; D0577A9D97E0CA92 CRC64;

Alignment Scores:
Pred. No.: 6.11 Length: 276
Score: 102.50 Matches: 50
Percent Similarity: 39.16% Conservative: 53

```



```
QY 328 CAAACATCTCTTCTAAATCTTAGAAGAGTTTGGAGAAATGGAT----- 375
Db 920 ProValAsnLeuLysArgilelleLysAsnAlaGlnLysGluLysGlyIleHisProSer 939
QY 376 -----GATGAAGAAGTGAAGAAATGTC 399
Db 940 IleGlyLysProSerAspMetAsnProIleTyrValIleGluLysIleLysLysLeu 959
QY 400 GATGAATTA---TCCATAAAACACGACGACAAATTTAAACAGACAGACCTTAAGCGCGCG 456
Db 960 AspAspLeuValIleLysGlyThrAspGluLeuSerThrGluAlaGlnThrAsnAla 979
QY 457 ACTTAAACGCTGTTTAAATGGGGGTTTAAATCTTATCAATAGCTGTCATGTTGCG 516
Db 980 ThrLeu---LeuPheSerMet----- 985
QY 517 AATGGCGTGCAGAAACCACTTCTAGGCGGCGTGTATCGCTTCGGGCAATCAGGTGCTT 576
Db 986 ---PheLeuArgSerThrPhePheLysGlyCysSerLeuLysAsnCysValLeuLeu 1004
QY 577 ACAAGAACCTGAGCTTTTAAACAGGTCCCTGTTGGCTGGATCAAT----- 621
Db 1005 LysLysProLeuThrPhe-----TipLeuValLysLeuGluLeuArg 1018
QY 622 -----ACAGGCGTATGGACGACGAGGATGATNT 648
Db 1019 PheHisAsnCysLeuAlaGlnProGlyGluMetValGlySerValAlaAlaGlnSerIle 1038
QY 649 GCAGGCGCGCTTATAGGTGAACATA 675
Db 1039 GlyGluProAlaThrGlnMetThrLeu 1047

RESULT 11
Q8IKG8 PRELIMINARY; PRT; 1455 AA.
AC Q8IKG8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Rhoptry protein, putative.
GN PF14_0637.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.B., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.;
RA "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; A5014826; AAN37250.1; -.
DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . . IEA.
DR InterPro; IPR002114; Hpr SerP.S.
DR PROSITE; PS00589; PTS HPR SER; 1.
SQ SEQUENCE 1455 AA; 174305 MW; 36FAAL760A898F99 CRC64;
```

Alignment Scores:  
Pred. No.: 7.97 Length: 1455  
Score: 101.90 Matches: 36  
Percent Similarity: 46.10% Conservative: 29  
Best Local Similarity: 25.53% Mismatches: 52  
Query Match: 7.49% Indels: 24

```
DB: 5 Gaps: 6
US-09-732-091-3 (1-759) x Q8IKG8 (1-1455)
QY 16 GATAGAGCTTGAATTTTAAAGCAATTGGAATCTAGTATTATTGGATTGTTTGAG 75
Db 333 GluArgGluLeuTyrTyrLeuGlyMetIleGlu-----GluLeuArgAsn 347
QY 76 GTGCTTTGTTTGTAAAGACGCGGAAAGACACATGACAAATGACACAGCTCCATA 135
Db 348 GluIleLysThrLysGluGluAsnGluGlyAsnAsnIleGluLysLeuGluAsnLysIle 367
QY 136 ----GAATACAAAGCATGCGCATGATTACGCTAAATACCGACAGAAATCGCTGAAGAG 192
Db 368 HisGluTyrGluLysGlnAsnGluGluLeuArgAsnGluLysGluLysLeuGlnSerThr 387
QY 193 TTGCAATACTATCGGACCAATAGTTTTCGCGAGTTTTCATTAAGCGGAGAGTCTTATAC 252
Db 388 IleAsnGluTyr---SerHisAsnPheAsnAsnLeuAsnAspHisAsnLysIleThrAsn 406
QY 253 AAAGAGATTTTATGCGATGTGCGATAATTAAGGTCAATTAACAAC-----AAG 303
Db 407 LysGlu-----CysGluGluLeuLysAsnAsnTyrAsnThrIleLysGlu 421
QY 304 AAACAGAAACGACTTTAATTGACAAACATCGCTTCTTAAATCTTAGAAGAGTTTG 363
Db 422 LysTyrGluArgLeuLysGluGluGlnGluIleTyrIleLys----- 435
QY 364 GAAGAAATCGATGATGAAGAAAGTGAAGAAATGTGCGATGAATATTCATCAATAAAACACG 423
Db 436 -----GluGluGluGluTyrLysSerLeuLeuAspGluLeuGluAsnAsn 452
QY 424 GAC 426
Db 453 Glu 453

RESULT 12
P87397 PRELIMINARY; PRT; 726 AA.
ID P87397;
AC P87397;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Heat shock protein hsp90.
OS Oncorhynchus tshawytscha (Chinook salmon) (King salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=74940;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=99262176; PubMed=10329464;
RX Palmsano A.N., Winton J.R., Dickhoff W.W.;
RA "Sequence features and phylogenetic analysis of the stress protein
RT hsp90alpha in chinook salmon (Oncorhynchus tshawytscha), a
RT poikilothermic vertebrate."
RL Biochem. Biophys. Res. Commun. 258:784-791(1999).
DR EMBL; U89945; AAB49983.1; -.
DR HSP; P07900; 1BYO.
DR GO; GO:0005254; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR001404; Hsp90.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00183; HSP90; 1.
DR PRINTS; PR00775; HEATSHOCK90.
DR SMART; SM00387; HATPase_c; 1.
DR PROSITE; PS00298; HSP90; 1.
SQ SEQUENCE 726 AA; 83505 MW; 1A2A962C471CBCE2 CRC64;
```

Alignment Scores:  
Pred. No.: 8.76 Length: 726  
Score: 100.50 Matches: 61

Percent Similarity: 35.46%      Conservative: 39  
Best Local Similarity: 21.63%      Mismatches: 81  
Query Match: 7.45%      Indels: 101  
DB: 13      Gaps: 11

US-09-732-091-3 (1-759) x P87397 (1-726)

```
QY 61 TTGAGATTGTTTGGAGTGCTGTTTGGTAAAGACGGCGGAAAAAGACACACATGAAAA 120
DB 414 MetAspLeuPheValGluLeuSerGluAspLysAspAsnTyrLysIlePheTyrGluGln 433
QY 121 CTGACCGCTCATAGATACAAAAGGCGATGCGATGATTCATCGCTAAATACACAGAAA 180
DB 434 PheSerLysAsnIleLysLeuGlyIleHisGluAsp-----AlaGlnAsnArgLysLys 451
QY 181 ATCGCTGAGAGTTCGATATCATGATGCGGACATAGTTTTCGCGATTTTCATTAAGCGNA 240
DB 452 LeuSerAspMetLeuArgTyrTyrThrSerAsnSerAsnAlaAsp-----Glu 467
QY 241 CGAGCTCTATACAAAGAGATTTCATGCGATGTGCGATAAATTAAGGTCAATTACAC 300
DB 468 MetValSerLeuLysGluTyrValSerArgMetLysAspThrGlnLysHisIleTyrTyr 487
QY 301 AAGAAAACCTGAACGACT----- 318
DB 488 IleThrGlyGluThrLysGluGlnValAlaAsnSerPheValGluArgLeuArgLys 507
QY 319 -----TTAATTGAA----- 327
DB 508 AlaGlyLeuGluValIleTyrMetIleGluProIleAspGluTyrCysValGlnGlnLeu 527
QY 328 -----CAAAACATGCTTCTTAAATCTTAGAAGAAAGTTTGGAGAAATG 372
DB 528 LysGluTyrAspGlyLysAsnLeuValSer---ValThrLysGluGlyLeuGluLeuPro 546
QY 373 GATGATGAAGAGTGAAGAATGTGCGATGAATTCATCAATAAAACACGCGACAAATTA 432
DB 547 GluAspGluAspGluLysLysGlnGluLeuAsnThrLys---PheGluAsnLeu 565
QY 433 AACAGACAGCCTTAAGCGCGGACTTTAAGCTGTTTAAATGGGGGTTTAAATCT 492
DB 566 CysLys----- 567
QY 493 TATCAATTAGCTGTTCATTGTTGCGAATCGCGTCAAAACCATCTAGGCGGTGTTTA 552
DB 568 -----ThrMetLysAspIleLeuAspLysLysIle 577
QY 553 TCG-----CTTGGCGCAATCAGGTGCTTACAGAACTCTGAGCTTTTAAACAGGTCT 606
DB 578 GluLysValSerValSerAsnArgLeuValSerProCysCysIleValThrSerThr 597
QY 607 GTTGGCTGGATCATTTACA----- 624
DB 598 TyrGlyThrThrAlaAsnMetGluArgIleMetLysSerGlnAlaLeuArgAspAsnSer 617
QY 625 -----GGCGPATGACAGCG-----ATTGATATTGCGAGCGCGCTTATAGGCTA 669
DB 618 ThrMetGlyTyrMetThrAlaLysLysHisLeuGluIleAsn----- 631
QY 670 ACCATACGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
DB 632 -----ProThrHisProIleValGluThrLeuArgGluLysAlaGluAlaAspLysAsn 649
QY 730 GATAAG 735
DB 650 AspLys 651
```

RESULT 13

ID Q7VHMO      PRELIMINARY:      PRT:      556 AA.

AC Q7VHMO;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN H0946  
OS Helicobacter hepaticus.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=32025;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 51449 / 391;  
RX MEDLINE=22709201; PubMed=12810954;  
RA Suerbaum S., Josephans C., Sterzenbach T., Drescher B., Brandt P.,  
RA Bell M., Droege M., Farman B., Fisch H.-P., Ge Z., Hoerster A.,  
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,  
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;  
RT "The complete genome sequence of the carcinogenic bacterium  
Helicobacter hepaticus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).  
DR EMBL; AB017147; AAP77543.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 556 AA; 61984 MW; 0E48CF5376BAC48D CRC64;

Alignment Scores:

Pred. No.: 10.5      Length: 556  
Score: 99.50      Matches: 53  
Percent Similarity: 38.57%      Conservative: 33  
Best Local Similarity: 23.77%      Mismatches: 104  
Query Match: 7.38%      Indels: 33  
DB: 16      Gaps: 6

US-09-732-091-3 (1-759) x Q7VHMO (1-556)

```
QY 94 GAGCGCGAAGACACACATGAAAACCTGACCGCTCCATAGATACAAAAGGCATGGC 153
DB 199 AspAspTyrLysArgHisLysGluAsnLeuGluLysMetIleGluLysGlyLysAsn 218
QY 154 GATGATTACGCTAAATACGCAAGAAATCGCTGAAGAGTTGCAATACTATGCGACAA 213
DB 219 ProGlu-----AsnGlnAsn 223
QY 214 AGTTTGGAGTTTCATTAAAGCGCAAGAGTCTTATACAAAGAGATTTTATCGCATGTG 273
DB 224 AlaValGluLysAlaGlnLysAlaGlnLysAlaLeuGluLysLeuAsnAlaAsnMet 243
QY 274 TCGGATAATTAAGGTCAATTACCAAGAAACCTGAACACGCTTA----- 321
DB 244 CysAsnArgLeuMetCysGluAsnProArgIleThrAlaValAlaThrGlnSerIleVal 263
QY 322 -----ATTGAACAAACATGCTTCTTAAATCTTAAAGAGAGTTTGAAGAA 369
DB 264 AlaSerGlyHisIleAlaGlnAlaGlyPheSerAspAlaIleAlaAlaLeuAlaThr 283
QY 370 ATGGATGATGA-----GAGTGAAGAAATGTCGAT-----GAA 405
DB 284 LeuAlaAsnGlyValIleTrpGluValLysAspMetPheGluGlySerIleAspThrGlu 303
QY 406 TTATCCATA-----AAAAACACGCAATTTAAACAGACAGCTTAAGCGCGCGACTTTA 462
DB 304 ThrSerIleLeuLysArgIleLysArgLeuLysArgThrIleGluAlaPheGlnAla 323
QY 463 ACGCTGTTTAAATGGGGGTTTAAATCTTATCAATTAGCTGTGCTGCTGCTGCTGCTGCT 522
DB 324 ThrPheGlyArgGlyAlaGlyPheGlyAlaIleAspAlaAlaValGlyValGlyGln 343
QY 523 GTCGAAAACCATCTTAGCGCGTGGTTTATCGCTTGGCGCAATCAGGTGCTTACAGA 582
DB 344 IlePheArgSerIleAlaGlyLys---LeuLysLeuValTrpAspLysIleArgThrAla 362
QY 583 ACTCTGATCTTTTAAACAGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 642
DB 363 LeuLysSerIleTyrAsnGlyIleValSerTyrIleLysGlyGluValSerAsnLeuArg 382
QY 643 GATATTGAGGCGCGCTTATAGGTTAACCATAACCGCATGCTGCTGCTGCTGCTGCTGCT 702
```



```

Db 383 GluLeuLeuGlyllelleleuLysSerLeuPheSerAlaAlaTrpValSerThrLeu 402
QY 703 CGCCTAAAA 711
Db 403 AlaLeuGlu 405

RESULT 14
Q882L0
ID Q882L0 PRELIMINARY; PRT; 633 AA.
AC Q882L0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN PSPT02616.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., Desoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidse T.,
RA White C., Fraser C., Collmer A.;
RA "Complete sequence of Pseudomonas syringae."
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL: AE016865; AAO56120.1; -.
DR TIGR: PSPT02616; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0004871; F:signal transducer activity; IEA.
DR GO: GO:0005935; P:chemotaxis; IEA.
DR GO: GO:0007165; P:signal transduction; IEA.
DR InterPro: IPR004089; Cmtaxis_trans.
DR InterPro: IPR003660; HAMP.
DR Pfam: PF00672; HAMP; 1.
DR Pfam: PF00015; MCPsигнал; 1.
DR PROSITE: PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE: PS00885; HAMP; 1.
KW Complete proteome.
SQ SEQUENCE 633 AA; 68090 MW; D15A8BC00AC58637 CRC64;

Alignment Scores:
Pred. No.: 10.5 Length: 633
Score: 99.50 Matches: 69
Percent Similarity: 40.89% Conservative: 50
Best Local Similarity: 23.71% Mismatches: 87
Query Match: 7.38% Indels: 85
DB: 16 Gaps: 18

US-09-732-091-3 (1-759) x Q882L0 (1-633)

QY 7 TACAAATATGATGAGACTTGGATTTTAAAGCAATTTGAATCTAGTATTATTGGAT 66
Db 70 TyrAspAlaGluArgAlaSerAspTrpLeuLysAlaLeuGluArgLeuGluAlaHisVal 89
QY 67 TTGTTTTCAGGTGCTGTTTGTGTTTAAAGCGCGGCAACAAATGAAAACCTGACC 126
Db 90 LysTyrAlaGlnGluValPhe-----AspSerProLeuAsnIleProLeuVal 105
QY 127 AGCTCC-----ATGATACAAAGGCGATGCGTGAT----- 159
Db 106 AsnAlaAlaGluAlaLeuValGluTyrArgValHisTyrAspAsnLeuMetArgAla 125
QY 160 -----TACGCTAAATACGACAGAAAGATCGCTGAA 189
Db 126 ThrAlaAlaArgGluAlaThrArgGlyAlaPheGlyGlnTyrAlaAspAlaGlyAlaGlu 145
QY 190 GAGTTGCAATATGAGCAATAGTTTTCGAGAGTTTCAATTAAGGCGAAGGA----- 243

```

```

Db 146 AspLeuGlnLysLeu-----AsnAlaPheAlaArgSerAspGluGlySerAlaSerGln 163
QY 244 -----GTCCTTATACAAAGAGATTTTATGCGATGTCGTGCGAT 279
Db 164 ArgAspAlaIleValGlnAlaMetThrLeuPheGlnLysMetArgPheAspLeu----- 181
QY 280 ABAATTAAAGCTCAATTACAAAGAACTGAACGACTTTAATTGACAAACATGCTT 339
Db 182 ---ArgGlyTyrThrTyrSerLeuLysAlaGluAsnArgAlaProAlaGluAlaSerMet 200
QY 340 TCTAAATCTTAGAA-----AGAAGTTTCGAAGAAATGGATGATGAGAA-----GTG 387
Db 201 SerAlaValIleAsnPheValLysGlyLeuGlnGlyPheAspSerGlnSerAlaThrIle 220
QY 388 AAGAAATGTGCGATGAATTA---TCCATAAAACACCGACAAAT-----TTAAAC 435
Db 221 LysHisLeuValAspSerMetValSerTyrGlnAsnThrMetAsnGlnPheThrAlaAla 240
QY 436 AGACAGCCCTTAAGCGCGGCGACTTTAACCGCTGTTTAAATGGGGGT-----TTTAAA 489
Db 241 GlnAlaSerIleAspGlnAlaGlnAlaGlyIleThrLysValIleGlyValLeuPheLys 263
QY 490 TCTTAT---CAATTAGCTGTCTTGTTCGAATGCGGTC----- 525
Db 261 SerAlaAspGlnLeuSer-----AlaAsnGlnValSerLeuArgIleGluAspVal 277
QY 526 -----GCAAAACCAATCTAGGCGGTGTTTATCGCTTCGGGCAATCAGTGCTTACA 579
Db 278 AspGlnAlaArgThrLeuLeuSerValTrpLeu-----ValAlaAla 291
QY 580 AGAACTCTGAGCTTTTAAACAGGTCCTGTTGGTGGATCATTCACAGCGGTATGACACAG 639
Db 292 LeuIleMetSerPheLeu-----AlaAlaTrpValIleThrArgLeu----- 305
QY 640 ATTGATATTCAGCGCGGCTTATAGGGTTAACCATACCGCATGTCATGTGGTGGCCACT 699
Db 306 -----IleValGlyPro----- 732
QY 700 TTACGCTTAAACACACAGCAAGCAATGAGAT 732
Db 314 LeuLysLeuAlaGluArgValAlaAspGlyAsp 324

RESULT 15
Q81J22
ID Q81J22 PRELIMINARY; PRT; 202 AA.
AC Q81J22;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pf10_0048.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7.
RX MEDLINE=22355705; PubMed=22368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Partea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum".
RL Nature 419:498-511(2002).
DR EMBL: AE014829; AAN35246.1; -.
KW Hypothetical protein.
SQ SEQUENCE 202 AA; 24259 MW; 48FD22A7F21165A6 CRC64;

```

## Alignment Scores:

Pred. No.: 11.6 Length: 202  
 Score: 99.00 Matches: 46  
 Percent Similarity: 41.62% Conservative: 31  
 Best Local Similarity: 24.86% Mismatches: 50  
 Query Match: 7.34% Indels: 58  
 DB: 5 Gaps: 9

US-09-732-091-3 (1-759) x Q81JZ2 (1-202)

```

QY 1C AAATATGATGACAGACTGGAATTTTAAAGCAATTCGAATCTAGTGAATTTATGATTTG 69
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 32 LysTyrGluGluAsnIleLeuLeuLysLysLeuMetArgThrGluAsnIleLysPhe 51
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 70 TTTGAGTCTGCTGTTTGGTAAAGACGGCGGAAAAAGACACAAAT----- 114
   : : : : : : : : : : : : : : : : : : : : :
Db 52 -----AspGlnLysTyrAsnAlaIleAsnIleLeu 61
   : : : : : : : : : : : : : : : : : : : : :
QY 115 GAAAAAAGT-----ACGAGCTCCATAGATACAAAAGGCGATGGCATGATTACGCTAAA 168
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 62 GluLysLeuLysAspIleGluAspIleGlnTyrGluAspIleLeuAsnAspTyrAlaLys 81
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 169 -----TACGCGAGAAAGAAATCGCTCAAGAGTTGCAATACTAT 204
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 82 AsnPheLeuThrPheLeuLysGlySerPheGluLysValaGluGluIleGlnIle--- 100
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 205 GGGAGCATAGTTTTCGAGTTTCATTAAAGCGGAGGAGTCTTATACAAAGAGATTTTA 264
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 101 -----AsnAspLysIleSerPheArgAsnGlu-----LeuIleLysGlnAspLeu 116
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 265 TCGGATGTGTGGATAAATAAGGTCAATTACAAACAGAAA----- 306
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 117 AsnAsnLysLeuGlnAsnLeuAsnHisAsnTyrAsnLeuGluTyrIleAsnLeuAsnLys 136
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 307 -----ACTGAA-----ACGACTTTAATTGAACAAAACATGCTT 339
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 137 CysLeuGlyThrGluAspPheGluSerGlnLeuAsnAlaSerLeuLysPheAsnSerLeu 156
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 340 TCTAAATCTTAGAAGAGTTTGAAGAAATGGATGATGAACAGTGAAG----- 390
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 157 LysAspIleLeuSerLysLeuLeuLysGluMetLysAspTyrGluIleGlnTyrAsnAsn 176
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 391 -----GAAATGTGCGATGAATATCCATAAAAAAC 420
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 177 LeuSerAspThrAspLysPheLeuAspMetLysIleCysGluLeuGluAsnTyrLeuAsn 196
   |||||: : : : : |||||: : : : : |||||: : : : :
QY 421 ACGACAAATTTAAAC 435
   |||||: : : : : |||||: : : : : |||||: : : : :
Db 197 AsnThrAsnGlnAsn 201
   |||||: : : : : |||||: : : : : |||||: : : : :
  
```

Search completed: July 5, 2004, 03:06:06  
 Job time : 87.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 5, 2004, 02:53:47 ; Search time 16.5 Seconds  
(without alignments)  
4749.589 Million cell updates/sec

Title: US-09-732-091-3  
Perfect score: 1349  
Sequence: 1 atggcacaacaatatgatag.....cgttgcaaatagataatccatt 759

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09732091/runat\_02072004\_182407\_19046/app\_query.fasta\_1.903  
-DB=Issued Patents AA -QFMT=fastan -SUFFX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09732091.cgn\_1\_18 @runat\_02072004\_182407\_19046 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	364	27.0	273	4	US-09-252-991A-29841
2	346.5	25.7	321	4	US-09-252-991A-29840
3	247	18.3	258	4	US-09-489-039A-8852
4	103	7.6	739	4	US-09-543-681A-6437
5	101.5	7.5	173	2	US-08-658-639-12
6	101.5	7.5	173	3	US-08-944-604-12
7	99	7.3	222	3	US-08-944-604-16
8	99	7.3	274	4	US-09-114-001C-5279
9	95	7.0	2125	4	US-09-919-172-29
10	91.5	6.8	1086	4	US-09-543-681A-7696
11	89.5	6.6	281	4	US-09-198-452A-749
12	89.5	6.6	543	4	US-09-328-352-5845

13	89.5	6.6	546	4	US-09-345-236B-98	Sequence 98, Appl
14	89.5	6.6	546	4	US-09-345-236B-121	Sequence 121, App
15	89.5	6.6	560	4	US-09-446-301A-50	Sequence 50, Appl
16	88.5	6.6	552	4	US-09-446-301A-4	Sequence 4, Appl
17	88.5	6.6	552	4	US-09-099-932-4	Sequence 4, Appl
18	88	6.5	488	4	US-09-489-039A-13363	Sequence 13363, A
19	87.5	6.5	732	2	US-08-533-669A-18	Sequence 18, Appl
20	87.5	6.5	732	4	US-09-307-143-4	Sequence 4, Appl
21	87.5	6.5	732	4	US-09-183-861-18	Sequence 18, Appl
22	87.5	6.5	732	4	US-09-022-765-18	Sequence 18, Appl
23	87.5	6.5	732	4	US-09-551-974A-18	Sequence 18, Appl
24	87.5	6.5	732	4	US-09-565-501A-18	Sequence 18, Appl
25	87.5	6.5	732	4	US-09-639-206A-18	Sequence 18, Appl
26	87.5	6.5	732	4	US-09-874-923-18	Sequence 18, Appl
27	87	6.4	491	3	US-09-029-267-2	Sequence 2, Appl
28	86.5	6.4	331	4	US-09-134-001C-3626	Sequence 3626, Ap
29	84.5	6.3	528	4	US-08-637-670-27	Sequence 27, Appl
30	84.5	6.3	691	4	US-09-830-217-12	Sequence 12, Appl
31	84	6.2	492	4	US-09-107-532A-7062	Sequence 7062, Ap
32	83.5	6.2	264	2	US-08-719-697-8	Sequence 8, Appl
33	83.5	6.2	264	3	US-08-727-616A-8	Sequence 8, Appl
34	83.5	6.2	264	4	US-09-481-756-8	Sequence 8, Appl
35	83.5	6.2	286	1	US-07-721-775A-2	Sequence 2, Appl
36	83.5	6.2	286	1	US-08-339-658-2	Sequence 2, Appl
37	83.5	6.2	286	3	US-09-263-933-7	Sequence 7, Appl
38	83.5	6.2	286	3	US-09-263-933-14	Sequence 14, Appl
39	83.5	6.2	286	3	US-09-263-933-21	Sequence 21, Appl
40	83.5	6.2	286	4	US-09-025-769B-265	Sequence 265, App
41	83.5	6.2	286	4	US-09-025-769B-362	Sequence 362, App
42	83.5	6.2	286	4	US-09-919-901-7	Sequence 7, Appl
43	83.5	6.2	286	4	US-09-913-901-14	Sequence 14, Appl
44	83.5	6.2	286	4	US-09-919-901-21	Sequence 21, Appl
45	83.5	6.2	299	4	US-09-025-769B-285	Sequence 285, App

#### ALIGNMENTS

##### RESULT 1

US-09-252-991A-29841  
; Sequence 29841, Application US/09252991A  
; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 29841

LENGTH: 273

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29841

##### Alignment Scores:

Pred. No.: 4.77e-34 Length: 273  
Score: 364.00 Matches: 85  
Percent Similarity: 55.83% Conservative: 49  
Best Local Similarity: 35.42% Mismatches: 88  
Query Match: 26.98% Indels: 18  
DB: 4 Gaps: 4

US-09-732-091-3 (1-759) x US-09-252-991A-29841 (1-273)

Qy 28 GAATTTTAAAGCAATTCGATTCAGTATTATTCGATTGTTGAGTGTCTGTTT 87

Db 28 GlnLeuLeuGluArgValSerAsnAspLeuGluProLeuValGluTyrIleLeu--- 46

Qy	88	GGTAAAGACGGCGGAAAAAGACACAATAGTAAATAATGACACGCTCCATAGATACATAAAAGG	147
Db	47	-----LysAlaargThrGluSerLeuSerLysGlnValaspPheLysArg	61
Qy	148	CATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTCGAATACATCTATGGG	207
Db	62	TrpHisProGluHisArgArgTyAlaSerAlaLeuaspGluLeuArgLeuPheGly	81
Qy	208	AGCAATAGTTTGGAGTTTCATTAAGAGCGAAGAGTCTTATACAAAGAGATTTTATGC	267
Db	82	GlyAsnSerPheAlaAsnLeuTirPargLysSerGlyProSerTyrThrGluValValArg	101
Qy	268	GATCTGTGCGGATAAATTAAAGCTCAATTACACAAGAAMAATCGAAACGACTTTAATTGAA	327
Db	102	AspValAlaGlyLysLeuLysValLysGlyValGlySerMetGluLeuLeuGluLeuGlu	121
Qy	328	CAAAACATGCTTCTAAATCTTAGAAGAAGTTTGGAGAANAATGSGATGATGAAGAAGTG	387
Db	122	GluAlaMetValGlnSerIleLeuArgGlnAlaLeuGluLysSerSerGlyGluAspArg	141
Qy	388	AAAGAAATGTCGATGAATTA-----TCCATAAAACACGCGACAAT	429
Db	142	ArgGluLeuGluGluIleLeuArgGluAlaGlyLeuAspLysThrLysMetThrAlaLeu	161
Qy	430	TTAAACAGACAAGCCTTAAGCGCGCGACATTTTAACTGCTTTTAAATGCGGGGTTTAAA	489
Db	162	LeuAsnGlySerAlaLeuSerGly-----LeuValValProAlaValAlaArgMetIle	179
Qy	490	TCATTATCAATAGCTGTCATGTTGGCAATCGCGTGCGCAAAACCATCTTAGGGCGTGT	549
Db	180	LeuTyrArgThrSerThrValIleValAsnSerMetAlaGlnLeuLeuGlyHisGly	199
Qy	550	TTATCGCTTCGG-----GGCAATCAGGTGCTTACAAAGAACTCTGAGCTTTTA	597
Db	200	LeuArgSerAlaValAlaAlaGlyGlyThrPheAlaGlyArgAlaValAlaLeu	219
Qy	598	ACAGGTCCTGTTGCTGGATCATTATACAGGCGTATGGACAGCAATGATATTCAGGCGCG	657
Db	220	AlaGlyProValGlyTyrValIleAlaGlyValTyrThrAlaValAspLeuAlaGlyPro	239
Qy	658	GCTTATAGGCTAACCAACACCGCATGCTATGTTGGTTCGCATTTAGCCTTAAACACAG	717
Db	240	AlaTyrArgValThrIleProCysValLeuHisIleAlaMetLeuArgLeuLysAlaArg	259

RESULT 2

US-09-252-991A-29840

; Sequence 29840, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29840

; LENGTH: 321

; TYPE: PR1

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29840

Alignment Scores:

Pred. No.:	5,79e-32	Length:	321
Score:	346.50	Matches:	82
Percent Similarity:	51.88%	Conservative:	42
Best Local Similarity:	34.31%	Mismatches:	112
Query Match:	25.69%	Indels:	3
DB:	4	Gaps:	3



;; TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF  
;; TITLE OF INVENTION: BREAST CANCER  
;; NUMBER OF SEQUENCES: 14  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Testa, Hurwitz & Thibault  
;; STREET: 125 High St.  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02110  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/658,639  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: MEYERS, THOMAS C  
;; REGISTRATION NUMBER: 36,989  
;; REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 248-7000  
;; TELEFAX: (617) 248-7100  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 173 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-658-639-12

Db 125 erLeuThrAspGluLeuSerAsnLeuProSerThrGlySerLeuSerVal---AlaA 144  
QY 476 TGGGGGGTTTAAATCTTATCAATAGTCATGTTGGGAATGGCGTCGCA 528  
Db 144 laGlyGlyLysLysAlaGluAlaAlaSerAlaLeuAlaAspAla 161

## RESULT 6

US-08-944-604-12  
; Sequence 12, Application US/08944604  
; Patent No. 6218131  
; GENERAL INFORMATION:  
; APPLICANT: KEESE, SUSAN  
; APPLICANT: OBAR, ROBERT  
; APPLICANT: WU, YING-JYE  
; TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF  
; TITLE OF INVENTION: BREAST CANCER  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Testa, Hurwitz & Thibault  
; STREET: 125 High St.  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,604  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MEYERS, THOMAS C  
; REGISTRATION NUMBER: 36,989  
; REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 173 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-944-604-12

Alignment Scores:  
Pred. No.: 0.00254 Length: 173  
Score: 101.50 Matches: 41  
Percent Similarity: 43.58% Conservative: 37  
Best Local Similarity: 22.91% Mismatches: 72  
Query Match: 7.52% Indels: 29  
DB: 2 Gaps: 5  
US-09-732-091-3 (1-759) x US-08-658-639-12 (1-173)  
QY 26 TGAATTTTAAAGCAATTCGAATCTAGTGAATTTATGATTTGTTGAGTGTCTTT 85  
Db 2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp----- 14  
QY 86 TTGGTAAAGACGGCGAAAAAGACACAATGAAAACTGCACGCTCCATAGATACAAA 145  
Db 15 -----GlnLysThrTrpValAlaProGlyTyrValArg 25  
QY 146 GGCATGGCGATGATTCAGCTAATACGACAGAAAGATCCCTGAGAGTTGCAATCATG 205  
Db 26 LysPheValLeuMetArgAlaAsnLleGlnAlaValSerLeuLys--lleGlnThrLeuL 45  
QY 206 GGAGC---AATAGTTTGGAGTTTTCATTAAAGGC-----GAAGGAGTCTTAT 250  
Db 45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65  
QY 251 ACAAGAGATTTTATGCGATGTGCGATAAATTTAAAGGTCAATTAACAAGCAAGAACTG 310  
Db 65 shArgGlnLeuLysLeuProGlnlleGlnLyslleMetMetGluPheGluArgGlnAlaG 85  
QY 311 AAACGACTTTAAATGCAAAAACATGCTTTCTTAAATCTTTAGAAAGAGTTTGGAAAGAA 370  
Db 85 luileMetAspMetLysGluGluMetMetAsnAspAlaileAspAspAlaMetGlyAspG 105  
QY 371 TGCATCATGAGNA-----GTGAAGAAATGTCGATGATGATTTATCCATAA 415  
Db 105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeu 125  
QY 416 AAACACGCGACAATTTAAACAGACAAAGCCTTAAGCGCGGAGCTTTTAAACGCTTTTAAA 475

Alignment Scores:  
Pred. No.: 0.00254 Length: 173  
Score: 101.50 Matches: 41  
Percent Similarity: 43.58% Conservative: 37  
Best Local Similarity: 22.91% Mismatches: 72  
Query Match: 7.52% Indels: 29  
DB: 2 Gaps: 5  
US-08-944-604-12

US-09-732-091-3 (1-759) x US-08-658-639-12 (1-173)  
QY 26 TGAATTTTAAAGCAATTCGAATCTAGTGAATTTATGATTTGTTGAGTGTCTTT 85  
Db 2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp----- 14  
QY 86 TTGGTAAAGACGGCGAAAAAGACACAATGAAAACTGCACGCTCCATAGATACAAA 145  
Db 15 -----GlnLysThrTrpValAlaProGlyTyrValArg 25  
QY 146 GGCATGGCGATGATTCAGCTAATACGACAGAAAGATCCCTGAGAGTTGCAATCATG 205  
Db 26 LysPheValLeuMetArgAlaAsnLleGlnAlaValSerLeuLys--lleGlnThrLeuL 45  
QY 206 GGAGC---AATAGTTTGGAGTTTTCATTAAAGGC-----GAAGGAGTCTTAT 250  
Db 45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65  
QY 251 ACAAGAGATTTTATGCGATGTGCGATAAATTTAAAGGTCAATTAACAAGCAAGAACTG 310  
Db 65 shArgGlnLeuLysLeuProGlnlleGlnLyslleMetMetGluPheGluArgGlnAlaG 85  
QY 311 AAACGACTTTAAATGCAAAAACATGCTTTCTTAAATCTTTAGAAAGAGTTTGGAAAGAA 370  
Db 85 luileMetAspMetLysGluGluMetMetAsnAspAlaileAspAspAlaMetGlyAspG 105  
QY 371 TGCATCATGAGNA-----GTGAAGAAATGTCGATGATGATTTATCCATAA 415  
Db 105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeu 125  
QY 416 AAACACGCGACAATTTAAACAGACAAAGCCTTAAGCGCGGAGCTTTTAAACGCTTTTAAA 475

Alignment Scores:  
Pred. No.: 0.00254 Length: 173  
Score: 101.50 Matches: 41  
Percent Similarity: 43.58% Conservative: 37  
Best Local Similarity: 22.91% Mismatches: 72  
Query Match: 7.52% Indels: 29  
DB: 2 Gaps: 5  
US-08-944-604-12

US-09-732-091-3 (1-759) x US-08-658-639-12 (1-173)  
QY 26 TGAATTTTAAAGCAATTCGAATCTAGTGAATTTATGATTTGTTGAGTGTCTTT 85  
Db 2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp----- 14  
QY 86 TTGGTAAAGACGGCGAAAAAGACACAATGAAAACTGCACGCTCCATAGATACAAA 145  
Db 15 -----GlnLysThrTrpValAlaProGlyTyrValArg 25  
QY 146 GGCATGGCGATGATTCAGCTAATACGACAGAAAGATCCCTGAGAGTTGCAATCATG 205  
Db 26 LysPheValLeuMetArgAlaAsnLleGlnAlaValSerLeuLys--lleGlnThrLeuL 45  
QY 206 GGAGC---AATAGTTTGGAGTTTTCATTAAAGGC-----GAAGGAGTCTTAT 250  
Db 45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65  
QY 251 ACAAGAGATTTTATGCGATGTGCGATAAATTTAAAGGTCAATTAACAAGCAAGAACTG 310  
Db 65 shArgGlnLeuLysLeuProGlnlleGlnLyslleMetMetGluPheGluArgGlnAlaG 85  
QY 311 AAACGACTTTAAATGCAAAAACATGCTTTCTTAAATCTTTAGAAAGAGTTTGGAAAGAA 370  
Db 85 luileMetAspMetLysGluGluMetMetAsnAspAlaileAspAspAlaMetGlyAspG 105  
QY 371 TGCATCATGAGNA-----GTGAAGAAATGTCGATGATGATTTATCCATAA 415  
Db 105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeu 125  
QY 416 AAACACGCGACAATTTAAACAGACAAAGCCTTAAGCGCGGAGCTTTTAAACGCTTTTAAA 475

Db 45 ysSerAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65  
 QY 251 ACAAGAGATTTTATCGGATGTCGATTAATTAAGTCAATTAACAAGAACTG 310  
 Db 65 snArgGlnLysLeuProGlnileGlnLysileMetGluPheGluArgGlnAlaG 85  
 QY 311 AAACGACTTTAATGAACAAACATCTTTCTAATAATCTTAGAAGAACTTTGGAAGAAA 370  
 Db 85 luileMeAspMetLysGluMetMetAsnAspAlaileAspAlaMetGlyAspG 105  
 QY 371 TGGATGATGACAAA-----GTGAAGAAATGTCGATGAATATCATCAATAA 415  
 Db 105 luGluAspGluGluGluSerAspAlaValValSerGlnValLysAspGluLys 125  
 QY 416 AAAACACGACATTTAAACACACAGCCCTTAAGCGCGGCGACTTTAAGCTGTTTAAAA 475  
 Db 125 eLeuThrAspGluLysSerAsnLeuProSerThrGlyGlySerLeuSerVal---AlaA 144  
 QY 476 TGGGGGGTTTAAATCTTATCAATAGCTGTCTGTCATGTCGGAATGCGTTCGCA 528  
 Db 144 laGlyGlyLysAlaGluAlaAlaAiaSerAlaLeuAlaAspAlaAspAla 161

## RESULT 7

US-08-944-604-16  
 ; Sequence 16, Application US/08944604  
 ; Patent No. 6218131  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KEESEE, SUSAN  
 ; APPLICANT: OBAR, ROBERT  
 ; APPLICANT: WU, YING-JYE  
 ; TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF  
 ; BREAST CANCER  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Testa, Hurwitz & Thibault  
 ; STREET: 125 High St.  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02110  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/944,604  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MEYERS, THOMAS C  
 ; REGISTRATION NUMBER: 36,989  
 ; REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 248-7000  
 ; TELEFAX: (617) 248-7100  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 222 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-944-604-16

Alignment Scores:  
 Pred. No.: 0.00555 Length: 222  
 Score: 99.00 Matches: 39  
 Percent Similarity: 47.80% Conservative: 48  
 Best Local Similarity: 21.43% Mismatches: 83  
 Query Match: 7.34% Indels: 12  
 DB: 3 Gaps: 5

US-09-732-091-3 (1-759) x US-08-944-604-16 (1-222)

QY 16 GATAGAGACTTGGAATTTTAAAGCAATTCGAATCTAGTGAATTTATTTGGAATTTTGGAG 75  
 Db 30 AsparGluArgGlnLysLeuGluThrGlnGlnLysLysileAlaAspLysLysLys 49  
 QY 76 GTCTTTGTTTTCGTAAA--GACGGCAAAAAGACACAAATGAAAACGTGACCAGCTCC 132  
 Db 50 MetAlaLysGlnGlnMetAspAlaValArgGlnMetAlaLysAspLeuValArgThr 69  
 QY 133 ATAGATACAAAGCGCATCGCGATGATTACGCTAAATACGCAGAAAGAAATCGCTGAAGAG 192  
 Db 70 ArgArgTyrValArgLysPheValLeuMetArgAlaAsnileGlnAlaValSerLeuLys 83  
 QY 193 TTGCAATCTATCGGAGC---AATAGTTTTCGAGTTTCAATTAAAGC----- 237  
 Db 90 IleGlnThrLeuLysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAla 109  
 QY 238 GAAGGAGTCTTATACAAAGAGATTTTATCGATGTGTGCGATAAATTAAGGTCAATTAC 297  
 Db 110 MetGlyThrMetAsnArgGlnLeuLysLeuProGlnileGlnLysileMetMetGluPhe 129  
 QY 298 AACAGAAACTGAAACGACTTTTAATTAACCAAAACATGCTTTCTAAAATCTTAGAAGA 357  
 Db 130 GluArgGlnAlaGluileMetAspMetLysGluGluMetMetAsnAspAlaileAspAsp 149  
 QY 358 AGTTTGGAGAAATGATGATGAAGAA-----GTGAAGAAATGTGCGAT 402  
 Db 150 AlaMetGlyAspGluGluAspGluGluSerAspAlaValValSerGlnValLeuAsp 169  
 QY 403 GAATTAATCCATAAAAAACACGACAACTTTAAACAGACAAAGCCTTTAAGCGCGGACTTTA 462  
 Db 170 GluLeuGlyLeuSerLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeu 189  
 QY 463 ACGCTGTTTAAATGCGGGTTTAAATCTTATCAATTAGCTGTCTCATTGTGGAATCGG 522  
 Db 190 SerVal---AlaAlaGlyGlyLysLysAlaGluAlaAlaAlaSerAlaLeuAlaAspAla 208  
 QY 523 GTCGCA 528  
 Db 209 AspAla 210

## RESULT 8

US-09-134-001C-5279  
 ; Sequence 5279, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI  
 ; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5279  
 ; LENGTH: 274  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-5279

Alignment Scores:  
 Pred. No.: 0.00608 Length: 274  
 Score: 99.00 Matches: 62  
 Percent Similarity: 40.77% Conservative: 44  
 Best Local Similarity: 23.85% Mismatches: 88  
 Query Match: 7.34% Indels: 66  
 DB: 4 Gaps: 14

US-09-732-091-3 (1-759) x US-09-134-001C-5279 (1-274)







; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-039A  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5845  
; LENGTH: 543  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
LS-09-328-352-5845

Alignment Scores:  
Pred. No.: 0.106 Length: 543  
Score: 89.50 Matches: 40  
Percent Similarity: 37.57% Conservative: 31  
Best Local Similarity: 21.05% Mismatches: 52  
Query Match: 6.63% Indels: 67  
DB: 4 Gaps: 10

US-09-732-091-3 (1-759) x US-09-328-352-5845 (1-543)

```
QY 16 GATAGACCTGGAATTTTAAAGCAATGGAATCTAGTATTGATTTGGATTGTTGAG 75
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 366 AspLysSerLeuGluGluLeuGluGluGluGluGluGluGluGluGluGluGlu 385
QY 76 GTGCTGTTGTTTGGTAAAGACGGCGGAAAAGACACAAATGAAAACCTGACCGCTCCATA 135
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 385 ----- 385
QY 136 GAATACAAAAGCGATGCGGAT-----GATTACGCT 165
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 386 GluTyrLysAsnLysGlyGlnIleLysPheValThrPheHisGlnSerPheSerGlyGlu 405
QY 166 AATACGCGAAGAAATC-----GCTGAAGAGTGTGCAATGCTATGAGGCAATAGTTTSGC 222
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 406 AspPheValGluGluIleArgAlaGluThrValGluTyrSerAspGlyLysLysAsnIle 425
QY 223 AGTTTCATTAAAGCGGAGGAGCTTATACAAAGAGATTTTATCGGAT----- 270
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 426 GluTyrProValValSerGlyValPhe-----LysLeuLeuCysAspThrAlaGlnSer 443
QY 271 -----GTGTCGATATAATTAAGGTCAATTAACAAAGAACTGAAACGACTTTAATT 324
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 444 LysValIleLeuGluSerGlnLysIleAsnPheAspSerAsnThr----- 458
QY 325 GACAAACATGCTTCTTAAATCTTGAAGAGAGTTTGGAGAAATG-----GATGATGAA 381
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 459 -----AsnGluIleTrpLysMetSerLeuGlyArgAlaGlyGluAspGlu 473
QY 382 GAAGTGAAGAAATGTGC-----GATGAATTA 408
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 474 AspIlePheAspTyrCysIleLysAsnHisCysValLeuLeuGlyTrpGlyAspGluLeu 493
QY 409 TCATTAATAAACCGGACCAATTTAAACAGACAAAGCTTAAGCGCGGCGACTTTAAACGCTG 468
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 494 AspPheSerAsnAla-----ValAsnArgLysGlnIleGlu-----GluIle 507
QY 469 TTTAAATCGGGGTTTAAATCTTATCAA 498
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 508 MetAspLysAsnGlyTyrGluAlaTyrArg 517
```

RESULT 13

US-09-345-236B-98  
; Sequence 98, Application US/09345236B  
; Patent No. 6521454

GENERAL INFORMATION:

APPLICANT: Becnel, James J.

APPLICANT: Tokuo, Fukuda

APPLICANT: Moser, Bettina

APPLICANT: Cockburn, Andrew

APPLICANT: White, Susan E.

APPLICANT: Undeen, Albert H.

; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal

; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates  
; FILE REFERENCE: 21042.0004  
; CURRENT APPLICATION NUMBER: US/09/345,236B  
; CURRENT FILING DATE: 1999-06-30  
; NUMBER OF SEQ ID NOS: 148  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 98  
; LENGTH: 546  
; TYPE: PRT  
; ORGANISM: mosquito baculovirus  
US-09-345-236B-98

Alignment Scores:  
Pred. No.: 0.106 Length: 546  
Score: 89.50 Matches: 60  
Percent Similarity: 36.86% Conservative: 34  
Best Local Similarity: 23.53% Mismatches: 92  
Query Match: 6.63% Indels: 69  
DB: 4 Gaps: 10

US-09-732-091-3 (1-759) x US-09-345-236B-98 (1-546)

```
QY 67 TTGTTTCAGGTGCTGTTTGTAAAGACGGCGAAAAGACACAAATGAAA- 120
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 153 ValPheGluValValGlyGlyGlyProGluGlyAspTyrAlaAlaGlyGluGluAspGlu 172
QY 121 -----CTGACCGCTCCATAGATACAAAGCGATGCGGATGATTAGCT 165
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 173 ValSerArgAsnSerLeuAsnPheAspMetAlaSerGluValGlnSerThrAspAlaAla 192
QY 166 AATACGCGCAAA-----AGNATCGCTGAGAGTTCGCAATGCTATGAGGAGCAATAGT 216
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 193 LysValMetGluLeuPheSerAlaLeuSerGluGluGlnArgAsnValIleLeuAsnAsn 212
QY 217 TTTCGCGAGTTTCATTAAGCGGAGGAGTCTTATACAAAGAGATTTTATGCGATGCTGC 276
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 213 PheGlyAlaAlaProSerGlySerGlyThr----- 222
QY 277 GATAAATTAAAGGTCAATTAACAACAGAAACTGAAACGACTTTTAATTGAACAAACATG 336
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 223 -----ThrProThrSerAlaGlnProAspMet 232
QY 337 CTTTCTAAATCTTAGAAGAGTTTGGAGAAATGATGATGATGAGAGTCAAGAAATG 396
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 233 -----GluValGluAspValGluThrValGluLys 242
QY 397 TGCATGAATATATCCATAAAAAACACGCAATTTAAAC-----AGACAAGCC 444
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 243 ProGluAsnLeuAsnAspIleIleThrAspGlnLeuArgAspPheMetAlaGlnGluLeu 262
QY 445 TTAAGCGCGGCGACTTTAAGCTGTTTAAATAGGG-----GTTTTTAATCTTAT 495
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 263 LysLysAlaAlaGluAsnTyrValProLysTrpGlySerThrValGlyGluSerLysSer 282
QY 496 CAATTAGCTGCTATTGTTGGAAATGCGTCGCAAAACCACTTCTAGGCGCTGTTTATCG 555
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 283 AlaLeuAlaIleThrValAlaAspArgValSerArgSerPheMetTyrGluGlyArgIle 302
QY 556 CTTGCGGCGCAATCAGGTG-----CTTACAAGA 582
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 303 ValAspTyrAsnGlnValValLeuHisIleLeuAspAsnTyrAspGlnArgLeuGluGlu 322
QY 583 ACTCTGAGCTTTTAAACAGTCTCTGTTGGCTGGATCATTT----- 621
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 323 LeuLeuSerPheArgThrLysThr-----TyrIleAlaGluGlyValProHisAsp 340
QY 622 ACAGCGCTATGGACACGCGATTGATATTGCGAGCGCGCTTATAGGGTAACCATACCGCA 681
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 341 SerLysValHisAspTyrValAspLeuThr-----GlnTyrArgGluThrValProTyr 358
QY 682 TGCAATTGCTGTGCCACTTTTACGCTTAAACACACAGCAAGCCCAAT 726
   |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 359 SerIleAlaLeuAsnAsnLeuSerArgGlyValAspGlnAlaAsn 373
```

```

Db      303 ValAspTyrAsnGlnValValLeuHisIleLeuAspAsnTyrAspGlnArgLeuGluGlu 320
QY      583 ACTCTGAGCTTTTAAACAGGTCCTCTGCTGGATCAATT-----TyrIleIleAlaGluGlyValProHisAsp 340
Db      323 LeuLeuSerPheArgThrLysThr-----TyrIleIleAlaGluGlyValProHisAsp 340
QY      622 ACAGCGCATGACAGCAGCATGATATTCAGGCGCGCTTATAGGTAACCATACCGGCA 681
Db      341 SerLysValHisAspTyrValAspLeuThr-----GlnTyrArgGluThrValProTyr 358
QY      682 TGCATTGTGCTGTCACCATTTAGCGCTTAAACACACAGCAAGCCAAT 726
Db      359 SerIleAlaLeuAsnAsnLeuSerArgGlyValAspGlnAlaAsn 373

RESULT 15
US-09-446-301A-50
; Sequence 50, Application US/09446301A
; Patent No. 6506893
; GENERAL INFORMATION:
; APPLICANT: EL SOLIH, NEVINE
; APPLICANT: ALIGNST, JERANIE
; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USE FOR DETECTING RESISTANCE
; TITLE OF INVENTION: TO STREPTOGRAMIN A OR TO STREPTOGRAMIN B AND RELATED
; TITLE OF INVENTION: COMPOUNDS
; FILE REFERENCE: 03715-0059
; CURRENT APPLICATION NUMBER: US/09/446,301A
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 560
; TYPE: PR1
; ORGANISM: Staphylococcus sp.
US-09-446-301A-50

Alignment Scores:
Pred. No.: 0.108 Length: 560
Score: 89.50 Matches: 37
Percent Similarity: 46.58% Conservative: 31
Best Local Similarity: 25.34% Mismatches: 47
Query Match: 6.63% Indels: 31
DB: 4 Gaps: 7

US-09-732-091-3 (1-759) x US-09-446-301A-50 (1-560)

QY      22 GACTTCGAATTTTAAAGCAATGGAACTAGTGATTTATTCGATTTGTTGAGTGCTT 81
Db      435 AspIleAspAlaValGluAlaLeuGlu-----GluLeuLeuIleThrTyrGluGlyVal 452
QY      82 GTTTTTGGTAAAGACGGCGGCAAAAAGACACAATGAAAACTGACAGCTCCATAGATAC 141
Db      453 ValLeuPheAlaSerHisAspLys----- 460
QY      142 AAAAGCATGGCGATGATTACGCTTAATACGAGAAAGAAATCGCTGAAGAGTTGCAATAC 201
Db      461 -----LysPheIleGlnAsnLeuAlaGluGlnLeuLeuIle 472
QY      202 TATGGAGCAATAGTTTTTCGAGTTTCATTAAGCGGAGGAGTCTTATACAAAGAGATT 261
Db      473 IleGluAsnAsnLysValLysLysPhe-----GluGlyThr---TyrIleGluTyr 488
QY      262 TTATCGCATGTGCGATAAATTAAGTCAATTACAAACAGAA-----ACTGAACG 315
Db      489 Leu---LysIleLysAspLysProLysLeuAsnThrAsnGluLysGluLeuLysGluLys 507
QY      316 ACTTTAATTGACAAAAATCGTTCTTAAATCTTAGAAAGAGTTTGGAGAAATGGAT 375
Db      508 LysMetIleLeuGluMetGlnIleSerSerLeuLeuSerLysIleSerMetGluGluAsn 527
QY      376 GATGAAGAAGTGAAGAAGATGGCGAT-----GAAATTATCCATAAAAAACACGCAAT 429
Db      528 GluGluLysAsnLysGluLeuAspGluLysTyrLysLeuLysGluLeuLysSer 547

```

Qy 430 TAAACAGACAGCCTTA 447  
|||||  
Db 548 LeuAsnLysAsnIleMet 553

Search completed: July 5, 2004, 03:07:23  
Job time : 21.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 5, 2004, 03:03:18 ; Search time 67.5 Seconds

(without alignments)  
7000.427 Million cell updates/sec

Title: US-09-732-091-3

Perfect score: 1349

Sequence: 1 atggcatacaaatatgatag.....cgttgcaaatagaatccatt 759

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 2553080

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTC\_spool/US09732091/tunat\_02072004\_182407\_19065/app\_query.fasta\_1.903  
-DB=Published Applications AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09732091@cgn\_1\_1\_35 -runat\_02072004\_182407\_19065  
-KCFU=6 -ICFU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-ZLONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
-----				

1	1279	94.8	253	9	US-09-732-091-4	Sequence 4, Appli
2	1279	94.8	265	9	US-09-732-091-44	Sequence 44, Appl
3	1270	94.1	253	12	US-10-335-977-9162	Sequence 9162, Ap
4	1270	94.1	253	12	US-10-335-977-9163	Sequence 9163, Ap
5	1270	94.1	256	12	US-10-335-977-9164	Sequence 9164, Ap
6	722	53.5	248	12	US-10-335-977-7699	Sequence 7699, Ap
7	716	53.1	237	12	US-10-335-977-7698	Sequence 7698, Ap
8	472.5	35.0	155	10	US-09-882-227-414	Sequence 414, Appl
9	252	18.7	49	9	US-09-732-091-20	Sequence 20, Appl
10	198	14.7	38	9	US-09-732-091-17	Sequence 17, Appl
11	194	14.4	41	9	US-09-732-091-19	Sequence 19, Appl
12	152	11.3	30	9	US-09-732-091-18	Sequence 18, Appl
13	150	11.1	30	9	US-09-732-091-16	Sequence 16, Appl
14	98.5	7.3	1009	12	US-10-282-122A-43832	Sequence 43832, A
15	97.5	7.2	815	16	US-10-437-963-191043	Sequence 191043, A
16	97	7.2	916	12	US-10-282-122A-76490	Sequence 76490, A
17	96.5	7.2	373	15	US-10-369-493-42	Sequence 42, Appl
18	96.5	7.2	701	14	US-10-032-585-7400	Sequence 7400, Ap
19	96.5	7.2	1102	14	US-10-156-761-14395	Sequence 14395, A
20	95	7.0	517	12	US-10-282-122A-54543	Sequence 54543, A
21	95	7.0	2125	9	US-09-919-172-29	Sequence 29, Appl
22	95	7.0	2649	14	US-10-205-219-169	Sequence 169, App
23	95	7.0	2649	15	US-10-341-434-220	Sequence 220, App
24	95	7.0	2649	15	US-10-341-434-230	Sequence 230, App
25	93.5	6.9	408	16	US-10-437-963-160944	Sequence 160944, A
26	93.5	6.9	1196	12	US-10-282-122A-52737	Sequence 52737, A
27	93	6.9	610	12	US-10-282-122A-47217	Sequence 47217, A
28	92.5	6.9	430	15	US-10-369-493-5229	Sequence 5229, Ap
29	92.5	6.9	996	9	US-09-815-242-5251	Sequence 5251, Ap
30	92.5	6.9	1009	9	US-09-815-242-12141	Sequence 12141, A
31	92	6.8	889	16	US-10-437-963-197045	Sequence 197045, A
32	91.5	6.8	241	12	US-10-412-699B-682	Sequence 682, App
33	91.5	6.8	1847	15	US-10-369-493-1075	Sequence 1075, Ap
34	90.5	6.7	659	12	US-10-424-539-222059	Sequence 222059, A
35	90.5	6.7	1163	15	US-10-452-024-107	Sequence 107, App
36	90	6.7	573	12	US-10-425-114-43207	Sequence 43207, A
37	90	6.7	744	16	US-10-437-963-122954	Sequence 122954, A
38	90	6.7	820	12	US-10-221-278-590	Sequence 590, App
39	90	6.7	820	15	US-10-291-172-590	Sequence 590, App
40	90	6.7	1089	14	US-10-032-585-7664	Sequence 7664, Ap
41	89.5	6.6	281	15	US-10-289-762-749	Sequence 749, App
42	89.5	6.6	460	12	US-10-425-114-63971	Sequence 63971, A
43	89.5	6.6	560	14	US-10-253-904-50	Sequence 50, Appl
44	89.5	6.6	869	12	US-10-282-122A-58561	Sequence 58561, A
45	89	6.6	384	12	US-10-260-708-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1  
US-09-732-091-4  
; Sequence 4, Application US/09732091  
; Patent No. US20020107368A1  
; GENERAL INFORMATION:  
; APPLICANT: Tian, Jing-Hui  
; APPLICANT: Walker, Richard I.  
; APPLICANT: Jackson, W. James  
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses  
; FILE REFERENCE: 7969-088  
; CURRENT APPLICATION NUMBER: US/09/732,091  
; CURRENT FILING DATE: 2000-12-07  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 253  
; TYPE: PRT  
; ORGANISM: Helicobacter sp.  
US-09-732-091-4

Alignment Scores:	7,22e-123	Length:	253
Pred. No.:	1279.00	Matches:	253
Score:			

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 94.81% Indels: 0  
 DB: 9 Gaps: 0

US-09-732-091-3 (1-759) x US-09-732-091-4 (1-253)

```

QY 1 ATGCGATACAAATATGATAGAGCTTCGAAATTTTAAAGCAATTTGGAATCTAGTATTTA 60
DB 1 MetAlaTyrlsTyAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20
QY 61 TTGGATTGTTTGGAGTCTGTTGTTTGGTAAAGACGGCGAAAAAGACACAATGAAAA 120
DB 21 LeuAspLeuPheGluValLeuValPheGlyAspGlyGluLysArgHisAsnGluLys 40
QY 121 CTGACGAGCTCCATAGATACAAAGCATGGCGATGATTCGCTTAATACGCGAAGAA 180
DB 41 LeuThrSerSerileGluTyrlsArgHisGlyAspAspTyrlsAlaLysTyrlsAlaGluArg 60
QY 181 ATCGCTGAAGAGTTGCAATATGAGCAATGTTTTCGAGTCTTTCATTAAGCGGAA 240
DB 61 IleAlaGluGluLeuGlnTyrlsGlySerAsnSerPheAlaSerPheileLysGlyGlu 80
QY 241 GGAGTCTTACAAAGAGTTTATCGATGTTGCGATTAATTAAGGTCAATTAACAAC 300
DB 81 GlyValLeuTyrlsGluLeuLeuCysAspValCysAspLysLeuLysValAsnTyrlsAsn 100
QY 301 AAGAAACCTGAAACGCTTAAATGACAAACATGCTTCTAAATCTTAGAAGAGT 360
DB 101 LysLysThrGluThrThrLeuLeuGlnAsnMetLeuSerLysileLeuGluArgSer 120
QY 361 TTGGAAGAAATGATGATGAAAGAGTGAAGAAATGCGATCAATTTATCCATAAAAAAC 420
DB 121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerileLysAsn 140
QY 421 AGCGACAATTTAAACAGACAGCTTAAGCGCGGACCTTAACGCTGTTTAAATGGGG 480
DB 141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160
QY 481 GGTGTTAAATCTTATCAATGCTGTCATTGTTGCGAATGCGGTGCGCAAAACCATCTTA 540
DB 161 GlyPheLysSerTyrlsGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrileLeu 180
QY 541 GGGGTGTTTATCGCTTCGGGCAATCAGGTGCTTTACAGAACTCTGAGCTTTTAAACA 600
DB 181 GlyArgGlyLeuSerileAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200
QY 601 GGTCTGTTGCTGATCATTACAGCGGTATGACAGCGATTCATTCAGCGCGCGCT 660
DB 201 GlyProValGlyTrpIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220
QY 661 TATAGGTTAACCATACCGCATGCTTGTGTTGCGACCTTTACGCTTAAACACACAGCAA 720
DB 221 TyrArgValThrileProAlaCysileValValAlaThrLeuArgLeuLysThrGlnGln 240
QY 721 GCCAATGAGATAGAAGTCTGTTGCAATAGAACATTCATT 759
DB 241 AlaAsnGlyAspLysSerLeuGlnIleGluSerile 253

```

## RESULT 2

US-09-732-091-44

Sequence 44 Application US/09732091

Patent No. US20020107368A1

GENERAL INFORMATION:

APPLICANT: Tian, Jing-Hui

APPLICANT: Walker, Richard L.

APPLICANT: Jackson, M. James

TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses

TITLE OF INVENTION: thereof

FILE REFERENCE: 7969-088

CURRENT APPLICATION NUMBER: US/09/732,091

CURRENT FILING DATE: 2000-12-07

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 44

LENGTH: 265

TYPE: PRT

ORGANISM: Helicobacter sp.

US-09-732-091-44

Alignment Scores:

Pred. No.: 7,31e-123 Length: 265  
 Score: 1279.00 Matches: 253  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 94.81% Indels: 0  
 DB: 9 Gaps: 0

US-09-732-091-3 (1-759) x US-09-732-091-44 (1-265)

```

QY 1 ATGCGATACAAATATGATAGAGCTTCGAAATTTTAAAGCAATTTGGAATCTAGTATTTA 60
DB 13 MetAlaTyrlsTyAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 32
QY 61 TTGGATTGTTTGGAGTCTGTTGTTTGGTAAAGACGGCGAAAAAGACACAATGAAAA 120
DB 33 LeuAspLeuPheGluValLeuValPheGlyAspGlyGluLysArgHisAsnGluLys 52
QY 121 CTGACGAGCTCCATAGATACAAAGCATGGCGATGATTCGCTTAATACGCGAAGAA 180
DB 53 LeuThrSerSerileGluTyrlsArgHisGlyAspAspTyrlsAlaLysTyrlsAlaGluArg 72
QY 181 ATCGCTGAAGAGTTGCAATATGAGCAATGTTTTCGAGTCTTTCATTAAGCGGAA 240
DB 73 IleAlaGluGluLeuGlnTyrlsGlySerAsnSerPheAlaSerPheileLysGlyGlu 92
QY 241 GGAGTCTTATACAAAGAGATTTTATCGATGTTGCGAATGCGAATTAATTAAGGTCAATTAACAAC 300
DB 93 GlyValLeuTyrlsGluLeuLeuCysAspValCysAspLysLeuLysValAsnTyrlsAsn 112
QY 301 AAGAAACCTGAAACGCTTAAATGACAAACATGCTTCTAAATCTTAGAAGAGT 360
DB 113 LysLysThrGluThrThrLeuLeuGlnAsnMetLeuSerLysileLeuGluArgSer 132
QY 361 TTGGAAGAAATGATGATGAAAGAGTGAAGAAATGCGATCAATTTATCCATAAAAAAC 420
DB 133 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerileLysAsn 152
QY 421 AGCGACAATTTAAACAGACAGCTTAAGCGCGGACCTTAACGCTGTTTAAATGGGG 480
DB 153 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 172
QY 481 GGTGTTAAATCTTATCAATGCTGTCATTGTTGCGAATGCGGTGCGCAAAACCATCTTA 540
DB 173 GlyPheLysSerTyrlsGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrileLeu 592
QY 541 GGGGTGTTTATCGCTTCGGGCAATCAGGTGCTTTACAGAACTCTGAGCTTTTAAACA 600
DB 193 GlyArgGlyLeuSerileAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 212
QY 601 GGTCTGTTGCTGATCATTACAGCGGTATGACAGCGATTCATTCAGCGCGCGCT 660
DB 213 GlyProValGlyTrpIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 232
QY 661 TATAGGTTAACCATACCGCATGCTTGTGTTGCGACCTTTACGCTTAAACACACAGCAA 720
DB 233 TyrArgValThrileProAlaCysileValValAlaThrLeuArgLeuLysThrGlnGln 252
QY 721 GCCAATGAGATAGAAGTCTGTTGCAATAGAACATTCATT 759
DB 253 AlaAsnGlyAspLysSerLeuGlnIleGluSerile 265

```

## RESULT 3

US-10-335-977-9162

Sequence 9162 Application US/10335977

Publication No. US20040052799A1

## GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

## NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

## COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997

## ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214

## INFORMATION FOR SEQ ID NO: 9162:

SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori

## FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...253  
SEQUENCE DESCRIPTION: SEQ ID NO: 9162:

US-10-335-977-9162

## Alignment Scores:

Pred. No.:	6,11e-122	Length:	253
Score:	1270.00	Matches:	251
Percent Similarity:	99.60%	Conservative:	1
Best Local Similarity:	99.21%	Mismatches:	1
Query Match:	94.14%	Indels:	0
DB:	12	Gaps:	0

US-09-732-091-3 (1-759) x US-10-335-977-9162 (1-253)

QY	1	ATGCATACAAATATGATAGACATTTGGAATTTTAAAGCAATTCGATCTAGTATTGATTA	60
Db	1	MetAlaTyrlYsTyAspAtgAspLeuGluPheLeuYsGlnLeuGluSerSerAspLeu	20
QY	61	TTGGATTGTTGGCTGCTGTTTGGTAAAGCGCGGAAAAAGACACAATCAAAAA	120
Db	21	LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys	40
QY	121	CTGACCAAGCTCCATAGATAACAAAGCATGGCGATGATACGTAATACCGAAGA	180
Db	41	LeuThrSerSerIleGluTyrlYsArgHisGlyAspAspTyrlAlaLysTyrlAlaG	60
QY	181	ATCGCTGAAGAGTTGCAATACTATGGAGCAATAGTTTTCGAGTTTCATTAAAGCGAA	240
Db	61	IleAlaGluGluLeuGlnTyrlYsGlySerAsnSerPheAlaSerPheIleLysGlyGlu	80
QY	241	GGAGTCTTATACAAAGAGATTTTATCGCATGTGCGATAAATTAAGGTCAATCAAC	300

Db	81	GlyValLeuTyrlYsGluIleLeuCysAspValCysAspLeuLysValAsnTyrlAsn	100
QY	301	AAGAAACTGAACGACTTTAATTCACAAACATGCTTTCTAAATCTTAAAGAACT	360
Db	101	LysLysThrGluThrLeuIleGlnAsnMetLeuSerLysIleLeuGluArgSer	120
QY	361	TTGGAAGAAATGGATGATGAAGAGTGAAGAAATGTCGGATGATTAATCCATAAAAC	420
Db	121	LeuGluGluMetAspAspGluValLysGluMetCysAspGluLeuSerIleLysAsn	140
QY	421	ACGACATTTAAACAGACAGCTTAAGCCGCGACTTTAAGCTGCTTTAAATGGGG	480
Db	141	ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly	160
QY	481	GGTTTAAATCTTATCAATAGCTGCTCATTTGCGATGCGGTGCGGCAAAACCATCTTA	540
Db	161	GlyPheLysSerTyrlGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu	180
QY	541	GGGCGTGTATCGTTGCGGCAATCAGGTGCTTACAAAGCACTCTGAGCTTTTAAACA	600
Db	181	GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr	200
QY	601	GSTCCTGTTGGCTGATCATTACAGGCGTATGACAGCGATTGATATTCAGGCGCGCT	660
Db	201	GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla	220
QY	661	TATAGGTAACCATACCGCATCGATTGCTGTTGCCACTTACGCTTAAAGCAACACACAA	720
Db	221	TyrlArgValThrIleProAlaCysIleValAlaThrLeuArgLeuLysThrGlnGln	240
QY	721	GCCATGGAGATAAGATCGTTGCAATAGATCCATT	759
Db	241	AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal	253

## RESULT 4

US-10-335-977-9162

Sequence 9162: Application US/10335977

Publication NO. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE &amp; COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

FILING DATE: 30-Dec-2002

APPLICATION NUMBER: US/10/335,977

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 9162:

SEQUENCE CHARACTERISTICS:

LENGTH: 253 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...253  
SEQUENCE DESCRIPTION: SEQ ID NO: 9163;  
US-10-335-977-9163

Alignment Scores:  
Pred. No.: 6.13e-122 Length: 253  
Score: 1270.00 Matches: 251  
Percent Similarity: 99.60% Conservative: 1  
Best Local Similarity: 99.21% Mismatches: 1  
Query Match: 94.14% Indels: 0  
DB: 12 Gaps: 0

US-09-732-091-3 (1-759) x US-10-335-977-9163 (1-253)

QY 1 ATGGCATCAAAATATGATAGAGACTTGGAAATTTTAAAGCAATTTGGAATCTAGTGATT 60  
Db 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20  
QY 61 TTGGATTGTTTGGAGTGCTGTGTTTGGTAAAGCGCGGAAAGACACAAATGAAAA 120  
Db 21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40  
QY 121 CTGACCAAGCTCCATAGATACAAAGCATGGCGCATGATTACGCTTAATACGCAAGAA 180  
Db 41 LeuThrSerSerLeuGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60  
QY 181 ATCGCTGAAGAGTTGCAATCTATGCGAGCAATAGTTTTCGAGTTTCATTAAAGCGCAA 240  
Db 61 IleAlaGluGluLeuGlnTyrLysArgHisAsnSerPheAlaSerPheLeuLysGlyGlu 80  
QY 241 GGAGTCTTATACAAAGAGATTTTATCGATGTGCGATTAATTAAGCTCAATTACAC 300  
Db 81 GlyValLeuTyrLysGluLeuLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100  
QY 301 AGAAACCTGAACGACTTTAATGACAAACATGCTTTCTAAATCTTAGAAGAAGT 360  
Db 101 LysLysThrGluThrThrLeuLeuGluGlnAsnMetLeuSerLysLysLeuGluArgSer 120  
QY 361 TTGGAAGAATGATGATGAGAGATCAAGAGATGCGCATGATGCAATATCCATAAAAC 420  
Db 121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerLysLysAsn 140  
QY 421 ACGGACAAATTTAAACAGACAGCGCTTAAGCGCGGACITTTAACGCTGTTTAAAAATGGGG 480  
Db 141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160  
QY 481 GCTTTTAAATCTTATCAATAGCTGTCATTTGCGAATGCGTTCGCAAAACCATCTTA 540  
Db 161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180  
QY 541 GGGCGTGTTTATCGCTTCGGGCAATCAGTGTCTTACAGACTCTGAGCTTTTAAACA 600  
Db 181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200  
QY 601 GGTCTCTGTTGGTGCATCATTAACAGCGTATGACAGCGATTGATATTGCGAGCGCGGCT 660  
Db 201 GlyProValGlyTyrPheIleThrGlyValTyrThrAlaIleAspIleAlaGlyProAla 220  
QY 661 TATAGGTAAACATACCGCATGCAATGTGTTGCCATTTAGCTTAAACAAACACAGCAA 720  
Db 221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240  
QY 721 GCCAATGGAGATAAGAGTCGTTGCAATAGATCCATT 759  
Db 241 AlaAsnGluAspLysSerLeuGlnIleGluSerVal 253

## RESULT 5

US-10-335-977-9164  
Sequence 9164 Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy B.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 9164:

SEQUENCE CHARACTERISTICS:

LENGTH: 256 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...256

SEQUENCE DESCRIPTION: SEQ ID NO: 9164;

US-10-335-977-9164

Alignment Scores:

Pred. No.: 6.13e-122 Length: 256

Score: 1270.00 Matches: 251

Percent Similarity: 99.60% Conservative: 1

Best Local Similarity: 99.21% Mismatches: 1

Query Match: 94.14% Indels: 0

DB: 12 Gaps: 0

US-09-732-091-3 (1-759) x US-10-335-977-9164 (1-256)

QY 1 ATGGCATCAAAATATGATAGAGACTTGGAAATTTTAAAGCAATTTGGAATCTAGTGATT 60  
Db 4 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 23  
QY 61 TTGGATTGTTTGGAGTGCTGTGTTTGGTAAAGCGCGGAAAGACACAAATGAAAA 120  
Db 24 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 43  
QY 121 CTGACCAAGCTCCATAGATACAAAGCATGCGCATGATTACGCTTAAATACGCAAGAA 180  
Db 44 LeuThrSerSerIleGluTyrLysArgHisGlyAspTyrAlaLysTyrAlaGluArg 63



QY	181	ATCGCTGAAGATTGCAATAC	TATGGAGCAATAGTTTTGCGAGTTTCATTAAAGCGCAA	240
Db	64	LEAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLeuGlyGlu	83	
QY	241	GGAGTCTTATACAAAGAGACTTTATGCGCATGTCGGATAAATTAAGTCAATTACAAAC	300	
Db	84	GlyValLeuTyrLysGluIleuGluCysAspValCysAspLysLeuLysValAsnTyrAsn	103	
QY	301	AGAAAACTGAACCGACTTTAATTGAACAAAAACATGCTTCTTAAAAATCTTGAAGAAAGT	360	
Db	104	LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer	123	
QY	361	TTGGAGAAANTGGATGATGAAGAAGTGAAGAAATGTGCGATGAATATCCATAAAAAAC	420	
Db	124	LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn	143	
QY	421	ACGGACAATTTAAACAGACAGCCTTAAGCGCGGCGACTTAAACGCTGTTAAAAATGGGG	480	
Db	144	ThrAspAsnLeuAsnArgGluAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly	163	
QY	481	GGTTTTAAATCTTATCAATTAGTGTCTCATTTGTGCGAATGGGTGCGCAAAAAACCATTTCTA	540	
Db	164	GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu	183	
QY	541	GGGCGTGGTTTTATGCTTCGCGGGCAATCAGTGCTTTACAGAACTCTGAGCTTTTTTAAACA	600	
Db	184	GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr	203	
QY	601	GGTCTCTGTGGCTCGATTCATTACAGGCGGTATGGACAGCGATGTGATTTGACGGCGCGCT	660	
Db	204	GlyProValGlyTyrIleThrGlyValTyrThrAlaIleAspIleAlaGlyProAla	223	
QY	661	TATAGGTAAACCATACCGCGCATGTGTGGTTCGCCACTTTACGCCCTAAAAAACACACCA	720	
Db	224	TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln	243	
QY	721	GCCAAATGGAGATGAAGTGTGTCGAATAGATCCATT	759	
Db	244	AlaAsnGluAspLysSerLeuGlnIleGluSerVal	256	

## RESULT 6

US-10-335-977-7699  
Sequence 7699, Application JS/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:

Db 231 ThrValProAlaCysValLeuValAlaThrLeuArgLysLys 244

## RESULT 7

US-10-335-977-7698  
; Sequence 7698, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10335,977  
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-Dec-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GYN-018

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7698:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 237 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori

FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...237  
SEQUENCE DESCRIPTION: SEQ ID NO: 7698:

US-10-335-977-7698

Alignment Scores:

Pred. No.: 7,79e-65 Length: 237

Score: 716.00 Matches: 140

Percent Similarity: 75.54% Conservative: 36

Best Local Similarity: 60.09% Mismatches: 53

Query Match: 53.08% Indels: 4

DB: 12 Gaps: 2

US-09-732-091-3 (1-759) x US-10-335-977-7698 (1-237)

QY 25 TTGGAATTTTAAAGCAATGGATCTAGTATTTTGGATTGTTGAGGTGCTGTT 84

Db 1 LeuGluPheLeuLysArgLeuSerSerAspLeuLysAspLeuPheAspAlaLeuVal 20

QY 85 TTGGAATAGACGGCGGAAAGAAACACATGAAAGTCCATGAGTCCATGAGTACAA 144

Db 21 TyrAspGluAspGlyThrLeuArgMetAsnGluLeuThrSerLeuThrGluTyrGln 40

QY 145 AGGCATGGCGATGATTACGCTAATACGAGCAAGATCGCTGAGAGGTGCAATCTAT 204

Db 41 ArgTyrGlyHisAspTyrAlaLysTyrProArgGluAlaGluLeuGlnArgTyr 60

QY 205 GGGAGCAATAGCTTTTGGAGTTTCATTAAAGCGAGGAGTCTTATACAAAGAGATTTTA 264

Db 61 GlyGlyAsnSerPheAlaAsnPhenPheArgAspGluGlyValLeuTyrLysGluLeuLeu 80

QY 265 TGGCATGTGCGATTAATTAAGTCAATTACAAAGAAAGTGAAGAACTGAAACGACTTAAT 324

Db 81 CysAspAlaCysAspHisLeuAspIleAsnTyrAsnGluArgSerAlaThrSerLeuIle 100

QY 325 GAACAAACATGCTTTCTAAATCTTAGAAAGAAAGTTTGGAGAAATCGATGATGAAGAA 384

Db 101 GluGlnAsnMetLeuSerLysLeuLeuLysAspSerLeuGluLysMetSerGlyArgGlu 120

QY 385 GTGAAGAAATGTCGATGAATTATCCATAAAACACGACCAATTTA-----AAC 435

Db 121 IleLysGluLeuCysAspGlyLeuGlyMetProAsnIleAspLysValIleGlyGluAsn 140

QY 436 AGACAAGCCTTAACGGCGGACCTTTAACGCTGTTTAAATGGGGGTTTAAATCTTAT 495

Db 141 LysGlnValLeuIleAlaSerValLeuThrLeuPheLysAlaGlyGlySerHisSerTyr 160

QY 496 CAATTAGCTGTCTATTGTCGATTCGCGTCCGAAACCAATCTAGGCGCTGGTTTA--- 552

Db 161 AlaLeuAlaValAlaValAlaAspAlaMetValArgGlnThrLeuGlyHisGlyLeuSer 180

QY 553 TCGCTTGGCGGCAATCAGGTCTTACAGAACTCTGAGCTTTTAAACAGGTCTGTGGC 612

Db 181 SerValValGlyLysValAlaLeuLysLysThrLeuAspIleLeuAlaGlyProIleGly 200

QY 613 TGGATCATTAACAGCGGTATGGACAGCGATGATTCGCGCTTATTCGCGCGCTTATAGGTAACC 672

Db 201 TrpValIleThrGlyAlaLeuValSerIleAsnLeuAlaGlyProAlaTyrArgValThr 220

QY 673 ATACCGCATGCTATTGTTGTCCTTTCACCTTTAGCCCTTAAATA 711

Db 221 ValProAlaCysValLeuValAlaThrLeuArgLysLys 233

RESULT 8

US-09-882-227-414

; Sequence 414, Application US/09882227

; Publication No. US20030158396A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Oomen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in t

; FILE REFERENCE: 06132/047002

; CURRENT APPLICATION NUMBER: US/09/882,227

; CURRENT FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 08/902,615

; PRIOR FILING DATE: 1997-07-29

; NUMBER OF SEQ ID NOS: 638

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 414

; LENGTH: 155

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-882-227-414

Alignment Scores:

Pred. No.: 8.85e-40 Length: 155

Score: 472.50 Matches: 93

Percent Similarity: 78.23% Conservative: 22

Best Local Similarity: 63.27% Mismatches: 29

Query Match: 35.03% Indels: 3

DB: 10 Gaps: 1

US-09-732-091-3 (1-759) x US-09-882-227-414 (1-155)

```

QY 112 AATGAAGAACTGACGAGTCCATAGATAACAAAGCGCATGCGATGATTACGCTAATAC 171
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 AsnGluAspLeuThrAsnSerThrGluTyrLysArgTyrGlyHisAspTyrAlaLysTyr 21
QY 172 CGAGAAAGAAATCGCTGAAGAGTTGCAATACTATGGAGCAATAGTTTTCGAGTTTCATT 231
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
22 ProArgA-GileAlaGluGluLeuGlnHisTyrGlyAsnSerPheAlaAsnPhePhe 41
QY 232 AAAGCGGAGGAGTCTTATACAAAGAGATTTATCGGATGTGTGGATPAAATTAAGGTC 291
Db :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 ArgAspGluGlyValLeuTyrLysGluLeuLeuCysAspAlaCysAspHisLeuLysVal 61
QY 292 AATTACAAACAGAAAACCTGAACGACTTTAATTGAACAAACATGCTTTCTAAATCTTA 351
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
62 AsnTyrAsnGluSerAlaThrSerLeuLileGluGlnAsnMetLeuSerLysLeuLeu 81
QY 352 GAAGAATTTGGAAGAAATGATGATGAAGAAGTGAAGAATGTCGATGAATATCC 411
Db :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 LysAspSerLeuGluLysMetSerArgArgGluLileLysGluLeuCysAsnGluLeuGly 101
QY 412 ATAAAAAACCGCAATTTA-----AACAGACAGCGCTTAAGCGCGCGACTTTA 462
Db :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 MetThrAsnIleAspLysValIleGlyLysAsnLysGlnValLeuIleAlaSerThrLeu 121
QY 463 ACGCTGTTTAAATCGGGGTTTAAATCTTATCAATTAGCTGTCTCATTTTCGCAATGCG 522
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 ThrLeuPheLysAlaGlySerHisSerTyrAlaLeuAlaValSerValAlaAspAla 141
QY 523 GTCGAAAACCACTCTAGGG 543
Db :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 MetValArgGlnThrLeuGly 148

RESULT 9
US-09-732-091-20
; Sequence 20, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-20
Alignment Scores:
Pred. No.: 3,55e-17 Length: 49
Score: 252.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.68% Indels: 0
DB: 9 Gaps: 0

US-09-732-091-3 (1-759) x US-09-732-091-20 (1-49)
QY 607 GTTGGCTGGATCATTACAGCGGTATGACGCGATTCATATTGCGAGCGCGCTTAGG 666
Db 1 ValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAlaTyrArg 20
QY 667 GTACCATACCGGATCGATTGCTGGCTTACGCTTAAACAAACACAGCAAGCAAT 726
Db 21 ValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGlnAlaAsn 40
QY 727 GGAGTAAAGAGTCGTTGCAATAGAA 753
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 41 GlyAspLysLysSerLeuGlnIleGlu 49
RESULT 10
US-09-732-091-17
; Sequence 17, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-17
Alignment Scores:
Pred. No.: 1,23e-11 Length: 38
Score: 198.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.68% Indels: 0
DB: 9 Gaps: 0

US-09-732-091-3 (1-759) x US-09-732-091-17 (1-38)
QY 157 GATTACGCTAAATACGACGAAAGAAATCGCTGAAGAGTTGCAATATCGGAGCAATAGT 216
Db 1 AspTyrAlaLysTyrAlaGluArgIleAlaGluGluLeuGlnTyrTyrGlySerAsnSer 20
QY 217 TTTCGGATTTTCATTAAGGCGGAGGAGTCTTATACAAAGAGATTTTATGCGAT 270
Db 21 PheAlaSerPheIleLysGlyGluGlyValLeuTyrLysGluIleLeuCysAsp 38

RESULT 11
US-09-732-091-19
; Sequence 19, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-19
Alignment Scores:
Pred. No.: 3,23e-11 Length: 41
Score: 194.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.38% Indels: 0
DB: 9 Gaps: 0

US-09-732-091-3 (1-759) x US-09-732-091-19 (1-41)
QY 433 AACAGACAAGCCTTAAGCGCGGCGACTTTAAACGCTGTTTAAATGCGGGGTTTAAATCT 492

```

Db 1 AsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGlyGlyPheLysSer 20  
QY 493 TRCAATAGTCGTCATTGTTGGAGTGGCGTCCGAAACCATTTCTAGGGCGTGGTTTA 552  
Db 21 TyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeuGlyArgGlyLeu 40  
QY 553 TCG 555  
Db 41 Ser 41

## RESULT 12

US-09-732-091-18  
; Sequence 18, Application US/09732091  
; Patent No. US20020107368A1

## ; GENERAL INFORMATION:

; APPLICANT: Tian, Jing-Hui  
; APPLICANT: Walker, Richard I.  
; APPLICANT: Jackson, W. James  
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses  
; TITLE OF INVENTION: thereof  
; FILE REFERENCE: 7969-088  
; CURRENT APPLICATION NUMBER: US/09/732,091  
; CURRENT FILING DATE: 2000-12-07  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 18  
; LENGTH: 33  
; TYPE: PRT  
; ORGANISM: Helicobacter sp.

US-09-732-091-18

## Alignment Scores:

Pred. No.: 6.37e-07 Length: 30  
Score: 152.00 Matches: 30  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.27% Indels: 0  
DB: 9 Gaps: 0

US-09-732-091-3 (1-759) x US-09-732-091-18 (1-30)

QY 361 TTGAAGAATGATGATGAAGAGTGAAGAATGTGCGATGAATTCATCAAAAAAC 420  
Db 1 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 20  
QY 421 ACGGCAATTTAAACAGACAGCTTACG 450  
Db 21 ThrAspAsnLeuAsrArgGlnAlaLeuSer 30

## RESULT 13

US-09-732-091-16  
; Sequence 16, Application US/09732091  
; Patent No. US20020107368A1

## ; GENERAL INFORMATION:

; APPLICANT: Tian, Jing-Hui  
; APPLICANT: Walker, Richard I.  
; APPLICANT: Jackson, W. James  
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses  
; TITLE OF INVENTION: thereof  
; FILE REFERENCE: 7969-088  
; CURRENT APPLICATION NUMBER: US/09/732,091  
; CURRENT FILING DATE: 2000-12-07  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 30  
; TYPE: PRT  
; ORGANISM: Helicobacter sp.

US-09-732-091-16

## Alignment Scores:

Pred. No.: 1.02e-06 Length: 30

Score: 150.00 Matches: 30  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.12% Indels: 0  
DB: 9 Gaps: 0

US-09-732-091-3 (1-759) x US-09-732-091-16 (1-30)

QY 1 ATGGCATAAAATATGATAGAGACTTGAATTTTAAAGCAATTCGAAATCTAGTCATTTA 60  
Db 1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20  
QY 61 TTGGATTTGTTGAGGTGCTTGTGTTTGGT 90  
Db 21 LeuAspLeuPheGluValLeuValPheGly 30

## RESULT 14

US-10-282-122A-43832

; Sequence 43832, Application US/10282122A  
; Publication No. US20040029129A1

## ; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wali, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 43832

; LENGTH: 1009

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-10-282-122A-43832

## Alignment Scores:

Pred. No.: 0.515 Length: 1009  
Score: 98.50 Matches: 66  
Percent Similarity: 37.88% Conservatives: 45  
Best Local Similarity: 22.53% Mismatches: 83  
Query Match: 7.30% Indels: 99  
DB: 12 Gaps: 16

US-09-732-091-3 (1-759) x US-10-282-122A-43832 (1-1009)

QY 13 TATGATAGAGCTTGAATTT-----TTAAGCAATTTGAA 48  
Db 708 TTAAGCAATTTGAAATTT-----TTAAGCAATTTGAA 48  
QY 49 TCTAGTATTTATTTGATTTGTTGAGGCTTTGTTTGGTAAAGACGCGCAAAAGA 108  
Db 728 LeuThrAspAsnAspLeuAsnGluThrIleAlaTrpArgGlyGluGlnGluGlu--- 746  
QY 109 CACAATGAAAACGACCAAGCTCCATGATCAATCAAAAGGCGATGCGTAAAT 168  
Db 747 TTAAGCAATTTGAAATTT-----TTAAGCAATTTGAA 48  
QY 169 TACGAGCAAGAAATCGCTGAAGATGCTGATCAATCAATGAGCAATGTTTGGAGTTC 228  
Db 757 ArgTyrHisGluPheGluMetGluIleAla-----ArgLeuGluSerLeu 771  
QY 229 ATTAAGCGCAAGGAGTCTTATACAAAGAGATTTATCGATGTGCGCATTAATTAAG 288  
Db 772 ThrLysAsp-----LysGluLeuLeuAspSer-----AspLysLeuLys 784  
QY 289 GTCAATTACCAACAGAAACTGAA-----ACGACTTTAATTTGAA 327  
Db 785 AspGluTyrGluGlnLysLysGluLysMetAsnThrLeuLeuAspGluTyrSerAlaVal 804  
QY 328 -----CMAAATGCTTCTTAATTTGAA 354  
Db 805 HisTyrGlnCysGlnAsnAsnLysAsnLysThrGlnSerIleValSerHisIleAsnTyr 824  
QY 355 AGAAGTTGGGAAGAAATGATCATCAAGAGTGAAGAAATGCGATCAATTTCCATA 414  
Db 825 LeuAsnGlnGluLeuLysAspGlnGlnGluIlePheGlnLeuAlaGluIleValSerGly 844  
QY 415 AAAACACGCAACATTTAAAC-----AGACAAAGCC 444  
Db 845 LysAsnAsnLysAsnLeuThrLeuGluAsnPheValLeuIleTyrTyrLeuAspGlnIle 864  
QY 445 TTAAGCGCGGCACTTTAAGCTGTTTAAAGTGGGGGTTTAAATCTTATCAATAGCT 504  
Db 865 IleAlaGlnAlaAsnLeuArgLeuAlaThrMetSerAsp-----AsnArgTyrGlnLeu--- 882  
QY 505 GTCATTGTTGCGAATGCGCTGCGCAAAACCATTTCTAGGCGGCTTTATCG----- 555  
Db 883 -----IleArgArgGluAlaValSerHisGlyLeuSerGlyLeuGlu 896  
QY 556 -----CTTGGCGGCAATCAGGCTTTAAGAACTCTGAGCTTTTAAACA 600  
Db 897 IleAspValPheAspLeuHisSerAsnLys-----SerArgHisIleSerSerLeuSer 914  
QY 601 GGTCTGTGTGGTATCATACAGGCGTATGACAGCGATTTGATTTGCGAGCGCGCT 660  
Db 915 Gly-----GlyGluThr 918  
QY 661 TATAGGTAAACATACCGGCTGATTTGTTGCTTACGCTTAAACACAGCAA 720  
Db 919 PheGlnSerSerLeuAlaLeuAlaLeuGlyLeuSerGluIle-----ValGlnGln 935  
QY 721 GCCAATGGAGATAGAGTCTTGAATATGATTCATT 759  
Db 936 GlnSerGly-----GlyIleSerLeuGluSerIle 945

## RESULT 15

US-10-437-963-191043  
; Sequence 131043, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21 (53221) B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 191043  
; LENGTH: 815  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_8739C.1.pap  
; US-10-437-963-191043

Alignment Scores:  
Pred. No.: 0 618 Length: 815  
Score: 97.50 Matches: 45  
Percent Similarity: 44.38% Conservative: 26  
Best Local Similarity: 28.12% Mismatches: 62  
Query Match: 7.23% Indels: 27  
DB: 16 Gaps: 8

US-09-732-091-3 (1-759) x US-10-437-963-191043 (1-815)

QY 4 GCATCAAAATATGATAGAGACTTTGAAATTTTAAAGCAATTTGAAATCTAGTATTATTG 63  
Db 531 AlaAsnArgMetAsnGlnAspLeu-----ValLeuGlnIleSerLysLeuGln 546  
QY 64 GATTGTTTGAAGTCTGTTGTTTGGTAAAGACGCGGCAAAAGACACAAATGAAACTG 123  
Db 547 AspGluPheAsnGluMetGlnGluGlyLeuThrAsnLysLeuGlyGluValGluSerVal 566  
QY 124 ACCAGCTCCATAGATACAAAGGATGCGGATGATTACGCTTAATACGACGAGAAATC 183  
Db 567 SerLysAlaLeu-----SerAspGluLeuValSerValLysGluMetVal 581  
QY 184 -----GCTGAAGAGTTGCAATATGAGCAATAGTTTTCGAGTTCATTAAA 234  
Db 582 HisLysGlyGlnGluGluLeuGlu---AlaThrSerAsnGlnLeuAlaSerIleValGlu 600  
QY 235 GCGGAGGAGTCTTATACAAAGAGATTTTATCGGATGCTGCGATAAATTAAGGTCAAT 294  
Db 601 AlaArgAspAsnLeuLysGlnLeuLeu---AspValPheLysLysLeu----- 616  
QY 295 TACAACAAGAAACTGAAACGACTTTAATTGAACAAACATCGCTTTCTAAATCTTAGAA 354  
Db 617 -----GluSerThrSerGlnGlnLeuValAspGluArgLysThrValThrThrLeuAsn 634  
QY 355 AGAATTTGGAA-----GAAATGGATGATGAAGAGTGAAGAGAAATG 396  
Db 635 ArgGluLeuGluAlaLeuValLysGlnLeuGlnMetAspSerGluAlaArgLysAlaLeu 654  
QY 397 TCGGATGAATTA-----TCCATAAAACACGCAATTTTAAACAGACAAAGCTTAAG 450  
Db 655 GluAlaAspLeuAspGluAlaThrLysSerLeuAspGluMetAsnArgSerAlaLeuSer 674

Search completed: July 5, 2004, 03:10:40  
Job time : 71.5 secs